

## SEARCH REQUEST FORM

Requestor's Name: \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Date: \_\_\_\_\_ Phone: \_\_\_\_\_ Art Unit: \_\_\_\_\_

### Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

### STAFF USE ONLY

Date completed: 04-16-03  
Searcher: Beverly E 4997  
Terminal time: 20  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: 23  
Number of Searches: \_\_\_\_\_  
Number of Databases: 1

#### Search Site

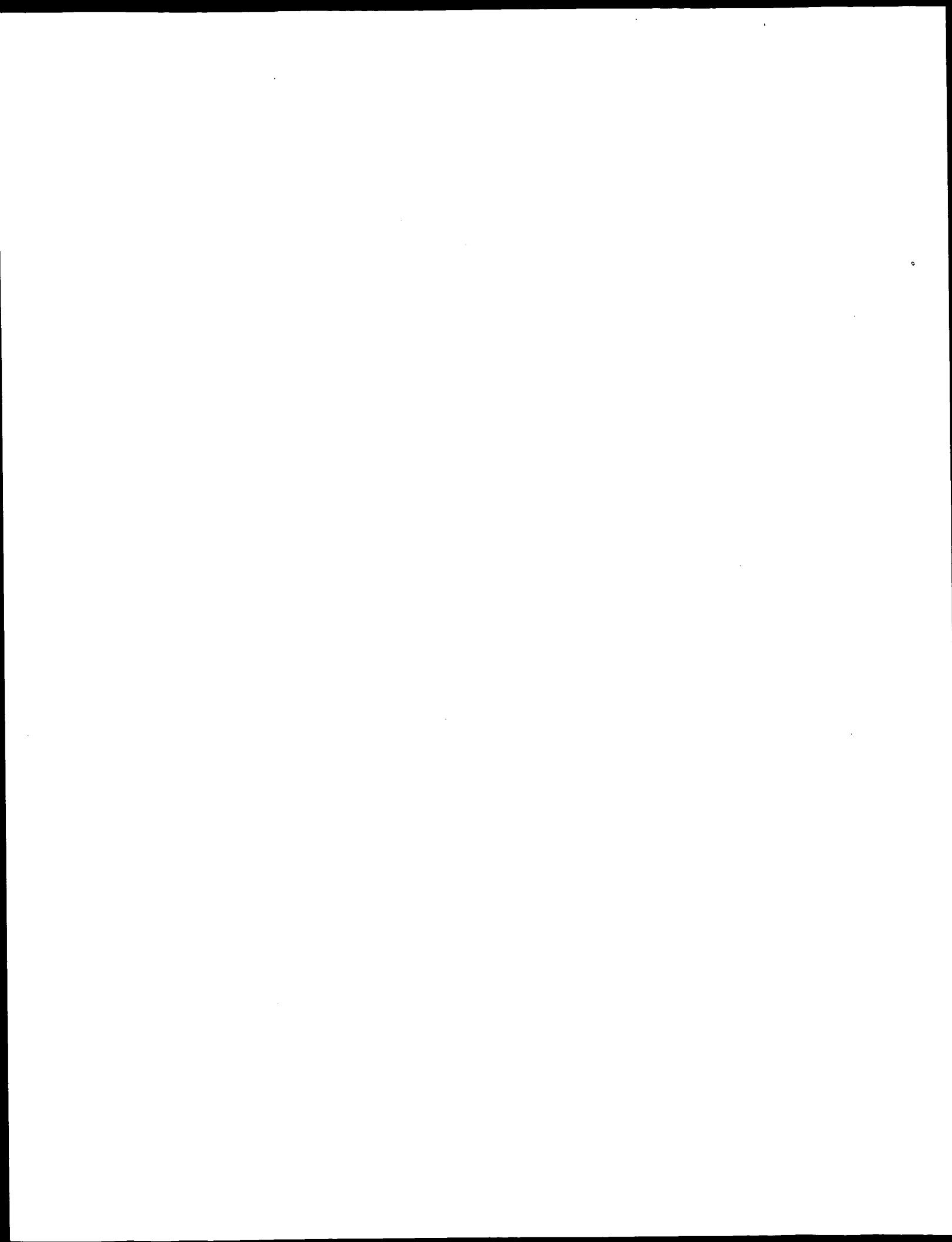
\_\_\_\_\_ STIC  
\_\_\_\_\_ CM-1  
\_\_\_\_\_ Pre-S

#### Type of Search

\_\_\_\_\_ N.A. Sequence  
\_\_\_\_\_ A.A. Sequence  
\_\_\_\_\_ Structure  
\_\_\_\_\_ Bibliographic

#### Vendors

\_\_\_\_\_ IG Suite  
\_\_\_\_\_ STN  
\_\_\_\_\_ Dialog  
\_\_\_\_\_ APS  
\_\_\_\_\_ Geninfo  
\_\_\_\_\_ SDC  
\_\_\_\_\_ DARC/Questel  
\_\_\_\_\_ Other CGN



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 13:06:01 ; Search time 3378 Seconds

(without alignments)  
9175.396 Million cell updates/sec

Title: US-09-905-173-23

Perfect score: 1065  
Sequence: 1 atgatacccccagagatttaa.....cccttcaagctctctttaa 1065

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_da: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pln: \*  
35: em\_htg\_rtd: \*  
36: em\_htg\_mam: \*  
37: em\_htg\_vrt: \*  
38: em\_sy: \*  
39: em\_htgo\_hum: \*  
40: em\_htgo\_mus: \*  
41: em\_htgo\_other: \*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1065	100.0	1065	6	AR043346
2	1065	100.0	1065	6	AR078122
3	1063.4	99.8	12497	1	AE000771
4	145.4	13.7	12075	1	AE013160
5	103.8	9.7	12452	1	AE010265
6	90.8	8.5	11932	1	AE002508
7	90.8	8.5	349980	6	AX044033
8	89.2	8.4	20334	1	AE001764
9	87.6	8.2	326301	1	NMA62491
10	86	8.1	22014	1	AE000963
11	80.6	7.6	188050	1	AL646072
12	72.2	6.8	1480	1	AF246314
13	68.8	6.5	11265	1	AE012885
14	64.2	6.0	1285	8	AT117255
15	64.2	6.0	1570	8	AY050832
16	61	5.7	7218	6	16494
17	60.2	5.7	17492	1	U67539
18	60	5.6	1242	8	NPL278767
19	56.8	5.3	1556	1	NHRA
20	55.2	5.2	300600	1	AE005369
21	54.4	5.1	13190	1	AE013011
22	46.2	4.3	1085	6	AX431883
23	45	4.2	301150	1	AE003190
24	44.2	4.2	25770	1	AE000964
25	44.2	4.2	25770	1	LU92974
26	43.6	4.1	26335	1	CJ11168X5
27	42.2	4.0	1092	6	AR078126
28	41.6	3.9	119184	1	D90909
29	40.8	3.8	10872	1	AE008031
30	40.8	3.8	11490	1	AE009065
31	40.6	3.8	189287	1	AE0068362
32	40.6	3.8	200918	9	AC012360
33	40.2	3.8	177585	2	AC094167
34	40.2	3.8	226608	2	AC105612
35	40	3.8	65785	2	AC101401
36	40	3.8	91365	2	AP004076
37	40	3.8	148159	2	AL772390
38	40	3.8	158050	8	AF480496
39	39.8	3.7	147565	2	AC112868
40	39.8	3.7	176744	2	AC128581
41	39.8	3.7	201420	2	AC099110
42	39	3.7	11080	1	AE006160
43	39	3.7	215296	2	AC093469
44	38.8	3.6	12257	1	AE009793
45	38.8	3.6	177014	2	AC022269

#### ALIGNMENTS

RESULT 1  
LOCUS AR043346 1065 bp DNA  
DEFINITION Sequence 23 from patent US 5814473.  
ACCESSION AR043346  
VERSION AR043346.1 GI:5964354  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1065)  
AUTHORS Warren, P. V. and Swanson, R. V.  
TITLE Transaminases and aminotransferases  
JOURNAL Patent: US 5814473-A 23 29-SEP-1998;  
FEATURES Location/Qualifiers

```

source          1. .1065
                /organism="unknown"
BASE COUNT     344 a      220 c      252 g      249 t
ORIGIN

```

BASE COUNT	344	a	220	c	252	g	249	t
ORIGIN								

Query Match	100.0%	Score 1065;	DB 6;	Length 1065;
Best Local Similarity	100.0%	Pred. No. 1,4e-290;		
Matches 1065; Conservative	0;	Mismatches	0;	Indels 0;
				Gaps 0

QY	1	ATTATACCCCGAGAGATTAAAGACCTTGAGAGCTTGACAAAGCGAGGGTCATCTCCGGCTCC	60
Db	1	ATGATATACCCCGAGAGATTAAAGACTTGAAAGCTTGACAAAGCGAGGGTCATCTCCGGCTCC	60
QY	61	GTCAGGCGTTTCCTCTAACGAAATTCCTCCCTAGACATTTCCCGAGGAGATAAAACAAGGCC	120
Db	61	GTCAGGCGTTTCCTCTAACGAAATTCCTCCCTAGACATTTCCCGAGGAGATAAAACAAGGCC	120
QY	121	TTAGAGAGATTTAAAAAAGGTTCCCTTGAAACAAATACCCAGACCCCGAGGAAAGGTTA	180
Db	121	TTAGAGAGATTTAAAAAAGGTTCCCTTGAAACAAATACCCAGACCCCGAGGAAAGGTTA	180
QY	181	AAAGGGGTTCTTGCGGATTTTTCGGGGTTAAGGAAAGAAATTTAGTCTCGGTTAACGCT	240
Db	181	AAAGGGGTTCTTGCGGATTTTTCGGGGTTAAGGAAAGAAATTTAGTCTCGGTTAACGCT	240
QY	241	TGCGAGCACTCATATACATACCTCTCATAGCTATTAGTGAACCTTACATACCCGTTTAC	300
Db	241	TGCGAGCACTCATATACATACCTCTCATAGCTATTAGTGAACCTTACATACCCGTTTAC	300
QY	301	ATACCTGTTCCTCCACCTTTCCCATGTACGAGTAAGTGGAAAGTTCTCGGAAAGCCCTC	360
Db	301	ATACCTGTTCCTCCACCTTTCCCATGTACGAGTAAGTGGAAAGTTCTCGGAAAGCCCTC	360
QY	361	GTAAGGTTTCAACTGAGAGCAAAACCTTGATTTAACACTTAGAAAGAAAGTATTGATTTATA	420
Db	361	GTAAGGTTTCAACTGAGAGCAAAACCTTGATTTAACACTTAGAAAGAAAGTATTGATTTATA	420
QY	421	GAGAAAGAAAAACCCGTTCTCGGGTACTTTGCTTACCCAAACAACCCACGGGAAACCTC	480
Db	421	GAGAAAGAAAAACCCGTTCTCGGGTACTTTGCTTACCCAAACAACCCACGGGAAACCTC	480
QY	481	TTTTCCAGGGAAAGATGAGAGATTAAGAAACGAGGGGTGTTTTCTGTAATGAGACAA	540
Db	481	TTTTCCAGGGAAAGATGAGAGATTAAGAAACGAGGGGTGTTTTCTGTAATGAGACAA	540
QY	541	GCCCTACTATCATTTACTCCCGGAAAAACCTTTCTGGAAGACGGCTCAAAAAGGAGATACG	600
Db	541	GCCCTACTATCATTTACTCCCGGAAAAACCTTTCTGGAAGACGGCTCAAAAAGGAGATACG	600
QY	601	GTAGTTTGGAGACACTTTCAAAAATCCGATATGGCGATTTAAAGGTAAGGATTTTATA	660
Db	601	GTAGTTTGGAGACACTTTCAAAAATCCGATATGGCGATTTAAAGGTAAGGATTTTATA	660
QY	661	GGGAAGGGGGAATGCTCGAATAATTAACAAGGGAACCTCCCTCAACAGTACCTAC	720
Db	661	GGGAAGGGGGAATGCTCGAATAATTAACAAGGGAACCTCCCTCAACAGTACCTAC	720
QY	721	CCCTCTCAGGTGATGGCAAAAGTTCTCTCAACGAGGAGAAAGATTTCTTATGAAAAG	780
Db	721	CCCTCTCAGGTGATGGCAAAAGTTCTCTCAACGAGGAGAAAGATTTCTTATGAAAAG	780
QY	781	ATACAGAGAGTTGTAACAGAGCGAGAAAGATGTACGACAAATGACAAAATGTAAAGGA	840
Db	781	ATACAGAGAGTTGTAACAGAGCGAGAAAGATGTACGACAAATGACAAAATGTAAAGGA	840
QY	841	GTTGAGGTTTTTCGAGTAAGGCTAACTTCTGCTTTTACGAACCCCTTACCCGCGCAC	900
Db	841	GTTGAGGTTTTTCGAGTAAGGCTAACTTCTGCTTTTACGAACCCCTTACCCGCGCAC	900
QY	901	GAGGTTTATCAGAGCTATCGAAAAGGGATGCTCGTCAGAGAAAGTTCTTACATGGAA	960
Db	901	GAGGTTTATCAGAGCTATCGAAAAGGGATGCTCGTCAGAGAAAGTTCTTACATGGAA	960

QY	1021	GAAGCACTGAGGAGGATATAAAATCCCTTTCAGCTCTCTTAA	1065
QY <td>961</td> <td>GAACTCCAAAAGAGCCTCAGGCTAACCTGAGGAAACCGGAAGAAACACAACTTCTG</td> <td>1020</td>	961	GAACTCCAAAAGAGCCTCAGGCTAACCTGAGGAAACCGGAAGAAACACAACTTCTG	1020
Db	961	GCACTCCAAAAGTCCCTCAGGCTAACGCTGATGGAAACCGGAAGAAACACAACTTCTG	1020
QY <td>1021</td> <td>GAACCACTGGAGAGGATATAAAATCCCTTTCAGCTCTCTTAA</td> <td>1065</td>	1021	GAACCACTGGAGAGGATATAAAATCCCTTTCAGCTCTCTTAA	1065

RESULT 2  
AR078122  
100701000  
1006 b-  
DMM  
J1100AR  
DATE 31-AUG-2000

LOCUS	AR078122	1065 bp
DEFINITION	Sequence 23 from patent US 5962283.	
ACCESSION	AR078122	
VERSION	AR078122.1	GI:10004868

KEYWORDS	Unknown.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 1065)
AUTHORS	Warren, P.V. and Swanson, R.V.
TITLE	Transaminases and Amnotransferases
JOURNAL	Patent: US 5962283-A 23 05-OCY-1999;
FEATURES	Location/Qualifiers
source	1..1065
BASE COUNT	344 a 220 c 252 g 249 t
ORIGIN	/organism="unknown"
Query Match	100.0%; Score 1065; DB 6; Length 1065;
Best Local Similarity	100.0%; Pred. No. 1,4e-290;
Matches 1065; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

OY		1	ATGTAACCCAGAGATTAAAGAAGCTTGAAGCTTTACAAGACGGAGGTCACTCCCGCCTCC	60
Db		1	ATGATACCCCGAGAGATTAAAGAACTTGAAAGCTTACAAGACGGAGGTCACTCCCGCCTCC	60
OY		61	GTCAGGCTTTCCTCTAAGCAATTCGCCCTACGACTTTCGAGAGATAAAACAABGGGC	120
Db		61	GTCAGGCTTTCCTCTAAGCAATTCGCCCTACGACTTTCGAGAGATAAAACAABGGGC	120
OY		121	TTAGAAGATTAATAAAGAGTTCCTTAACAATAACCACGCCGGAAGGAAGTTA	180
Db		121	TTAGAAGATTAATAAAGAGTTCCTTAACAATAACCACGCCGGAAGGAAGTTA	180
OY		181	AAGAGCGTTTCTGGGATTTTTCTGGCGTTAAGGAACAAATTTAGTTCTCGTTAACGGT	240
Db		181	AAGAGCGTTTCTGGGATTTTTCTGGCGTTAAGGAACAAATTTAGTTCTCGTTAACGGT	240
OY		241	TGGAAGCACTCATATACCTACCTCTCAATAGCTATAGTGAACTTACATACCCGTTTAC	300
Db		241	TGGAAGCACTCATATACCTACCTCTCAATAGCTATAGTGAACTTACATACCCGTTTAC	300
OY		301	ATACCTGTTCCCACCTTCCCATGTAGAGATTAAGTCGAAAGTCTCGGAACACCCCTC	360
Db		301	ATACCTGTTCCCACCTTCCCATGTAGAGATTAAGTCGAAAGTCTCGGAACACCCCTC	360
OY		361	GTAAGGTTCAACGTGAGCAAGAACTTGATATACCTTAGAAGAAAGTATTGATTAAATA	420
Db		361	GTAAGGTTCAACGTGAGCAAGAACTTGATATACCTTAGAAGAAAGTATTGATTAAATA	420
OY		421	GAGAAAGAAAAACCGTCTCGGACTTTGCTTACCCAACAACCCACAGGGGAAACCTC	480
Db		421	GAGAAAGAAAAACCGTCTCGGACTTTGCTTACCCAACAACCCACAGGGGAAACCTC	480
OY		481	TTTTCACAGGGGAAAGTTGAGAGATTAAGAAACAGGGGTGTTTTCTGTGTAATAGACGA	540
Db		481	TTTTCACAGGGGAAAGTTGAGAGATTAAGAAACAGGGGTGTTTTCTGTGTAATAGACGA	540
OY		541	GCCCTACTATCATTTACCGGAGAAACCTTTCTGGAAGACGCGCTCAAAAGGGAAGATACG	6000
Db		541	GCCCTACTATCATTTACCGGAGAAACCTTTCTGGAAGACGCGCTCAAAAGGGAAGATACG	6000

OY	601	GTAGTTTGGAGACACTTTCACAAAATCGGTATGCGAGTTTAAGGATGAGATTTTAATA	660
Db	601	GTAGTTTGGAGACACTTTCACAAAATCGGTATGCGAGTTTAAGGATGAGATTTTAATA	660
OY	661	GGGAAGGGGGAATATGCTCAGAAATTAAACAGGTAAGACTCCCTTCAACGTGACCTAC	720
Db	661	GGGAAGGGGGAATATGCTCAGAAATTAAACAGGTAAGACTCCCTTCAACGTGACCTAC	720
OY	721	CCCTCTCAGGTATGCGAAAAGTTCTCCTCAGGAGGGAAGGATTCCTATGGAAG	780
Db	721	CCCTCTCAGGTATGCGAAAAGTTCTCCTCAGGAGGGAAGGATTCCTATGGAAG	780
OY	781	ATACAGAGAGGTTTAAACAGAGCGAAGAGATGTACGACGAATGACAAATAGAGA	840
Db	781	ATACAGAGAGGTTTAAACAGAGCGAAGAGATGTACGACGAATGACAAATAGAGA	840
OY	841	GTTGAGGTTTTCAGAGTAAGGCTTCTTCTTTCAGAGAGCCCTTACCCGCCAC	900
Db	841	GTTGAGGTTTTCAGAGTAAGGCTTCTTCTTTCAGAGAGCCCTTACCCGCCAC	900
OY	901	GAGGTTTATCAGAGACTCTGTAAGAAAGGATGCTCGTCAGAGAGCTATCTACATGAA	960
Db	901	GAGGTTTATCAGAGACTCTGTAAGAAAGGATGCTCGTCAGAGAGCTATCTACATGAA	960
OY	961	GCAGTCCAAAAGTGCTCAGGCTAGCGTAGGGAACCGGAAGAAACACAGATTCTG	1020
Db	961	GCAGTCCAAAAGTGCTCAGGCTAGCGTAGGGAACCGGAAGAAACACAGATTCTG	1020
OY	1021	GAGCAGCTGGAGAGAGATTAATCCCTTTCAGAGCTTCTTTAA	1065
Db	1021	GAGCAGCTGGAGAGAGATTAATCCCTTTCAGAGCTTCTTTAA	1065
RESULT 3	AE000771/c	12497 bp	DNA linear BCT 25-MAR-1998
LOCUS	AE000771	12497 bp	DNA linear BCT 25-MAR-1998
DEFINITION	Aqulifex aeolicus section 103 of 109 of the complete genome.		
ACCESSION	AE000771	AE000657	
VERSION	AE000771.1	GI:2984286	
KEYWORDS			
SOURCE	Aqulifex aeolicus.		
ORGANISM	Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifices.		
REFERENCE	1 (bases 1 to 12497)		
AUTHORS	Decker, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Anjaj, M., Huber, R., Feldman, R. A., Short, J. M., Olson, G. J. and Swanson, R. V.		
TITLE	The complete genome of the hyperthermophilic bacterium Aqulifex aeolicus		
JOURNAL	Nature 392 (6674), 353-358 (1998)		
MEDLINE	98196666		
PUBMED	9537320		
REFERENCE	2 (bases 1 to 12497)		
AUTHORS	Decker, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Anjaj, M., Huber, R., Feldman, R. A., Short, J. M., Olson, G. J. and Swanson, R. V.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego, CA 92121		
COMMENT	Putative indicates no similarity to known proteins		
FEATURES	Hypothetical indicates similarity to a protein of unknown function.		
source	1. 12497		
gene	/organism="Aqulifex aeolicus"		
CDS	/strain="VF5"		
	/db_xref="taxon:63363"		
	40..510		
	/gene="aq_2066"		
	40..510		
	/gene="aq_2066"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
gene	/protein_id="AAC07821.1"		
CDS	/db_xref="GI:2984297"		
	/translation="MESLKETLEAEVEGYGIGIIFLIAFTESIOPVPPDPFIAGATA		
	EGLNPLISALVATIGSVLGIVAVYPIGRFGEVEVAKLGEKEFLKAEAFYRRYVMA		
	VITAGTLPPIPFKFCWLSGILEVEVRFTLASHIGRFPFPLFALFGHSIQGFL"		
	387..860		
	/gene="aq_2067"		
	387..860		
	/gene="aq_2067"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AAC07822.1"		
	/db_xref="GI:2984298"		
	/translation="MLAFGNTRTRSGEYVFGFLYRVQSVSEVPLCEVWTLNSGFPITKE		
	LMSFLVEVMTPLGKGESEYQVAVRVDIIDKSGLPYLTIPMGTTIEGDMDEVMEVL		
	KGFELAKEDCNISITIMKIDYRKKGRLIKVSVYQEKIORDININATVEKKE"		
	934..1752		
	/gene="argB"		
	/note="aq_2068"		
	934..1752		
	/gene="argB"		
	/codon_start=1		
	/transl_table=11		
	/product="acetylglutamate kinase"		
	/protein_id="AAC07812.1"		
	/db_xref="GI:2984288"		
	/translation="MIKYGSAHDEBELRESFARDVLLKYVINPVIYHGGPQISK		
	TLEKFGIKRPFVGMKRTDEETEMHYVEMVLIIDKIDYALINRYSGEKIYAVGSGR		
	DRLKLRKLRKERYSELGIPVEEDIGFVEIYDVEDELIETLSINFLIPVAPVG		
	VBEQGEAVYVNDLIAASEIAGEIKAKELIYLTDTGTGIDEKELISSSKRAEELIK		
	KGVIRGMIIPKRSALRALKKVKVHIIDGNVXKSHILEVFTKGVETLE"		
	1765..2796		
	/gene="obg"		
	/note="aq_2069"		
	1765..2796		
	/gene="obg"		
	/codon_start=1		
	/transl_table=11		
	/product="GTP-binding protein"		
	/protein_id="AAC07816.1"		
	/db_xref="GI:2984292"		
	/translation="MEKFDVRYKIFVKGKGGDGAVAFLRKRYRKGSPAGDGGKG		
	DVLIVATSSKHTLIDPKYKHYIAONGRPGKKHGGDEGLIITYVYGVVDDAGT		
	ADVGLDVIKREGSKCYVAGGKGRNAPFAPITNOAPYAEQGEGERKWTILEKLI		
	ADVGLVGPFPNAGKSTLSRLTKARKIADYDPTTLPNVGMELDWERLVIADIPGL		
	IEDAHKAGLGHFRLHIERTKFLAHVIDVSDFRERFVQAFEAINRELEIYSPKLAQ		
	KPOIVVANKIDALSDRSLSELEKYEKGEYFVAVSALTGEIELEKGMKYEET		
	RDESAQVT"		
	2822..3997		
	/gene="aq_2071"		
	2822..3997		
	/gene="aq_2071"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AAC07819.1"		
	/db_xref="GI:2984295"		
	/translation="MLQYRVKRGVEDRIRGFFPMVYRSEIASFSRKKRKEIIVVVDV		
	NGKFLGYVINDEVNIAIRILSFKNEVEPELIRKRIKQAYEYKRLIYISNAYRLV		
	HSEGLPLGLIVDYGSLAVEFTYVGMORREVLIDTLELPRKGIYEVNIEIAR		
	VGLDVEERVIYGEVEVITMEHDLKVINIPOGOKTGFLLDRNARKIPRNIVFG		
	DRCILDVFCHTGSGPALNMRAGAGVIGVDSIELAKKEKARNLNGTDLIKVYANAF		
	DYIRLDRKKEKRPYITIIDPSPKRNRAVNALRGKELCVRLKIAKSGCYIAITS		
	CSFHTEBHLIDVLQASTDVROVRYIAKTFDDLHPWIIQMPNTLYLGWVEV"		
	3994..5151		
	/gene="aq_2073"		
	3994..5151		
	/gene="aq_2073"		
	/codon_start=1		
	/transl_table=11		
	/product="putative protein"		
	/protein_id="AAC07818.1"		

/db\_xref="GI:2984294"  
 /translation="MKKHVALVSEGNIFOSLIKLVGGILTSGLSLIDGAIHSLSDA  
 TASLIATLSTKFEIKSERPPGLYKLENGALVLEFLFTAMEIIOQLKEVIN  
 FENLPIGIVTVLSLTLSELERAGKLSAPLIDASFTLIDASSFLVLSL  
 SSYFGINIRYAVAVALLIYVATFELKEQALIDISADKEVEIKRILISPE  
 VSEYKRLVNRAGRLFDIVATINTDDFLIDELISKAIDEIERKIMKEPNVDYIYE  
 PCCKGTSTAVAKODGVAFRFDYKIIIFPENEGNEPIETKGDDECIACKELAR  
 RNDIVISGTHPIPSNKAKWILKNNVWVETKEENPEALSEVSQIKENT"  
 5148..6395  
 /gene="murD"  
 /note="aq\_2075"  
 5148..6395  
 /gene="murD"  
 /codon\_start=1  
 /transl\_table=11  
 /product="UDP-N-acetylmutamoylalanine-D-glutamate ligase"  
 /protein\_id="AAC07814.1"  
 /db\_xref="GI:2984290"  
 /translation="MKFLVGLGRSGKALKLKERGEVYAGDDSQNPELMREVLGE  
 VDVIVSPGIPPSHPLMKELKREKEVYGELELAYRPFKGVIALITGDKSTTRLT  
 YLIKKPPEVPEAGNIGKPSPEVULENPGIALVESSQGTIKTFRFNIGATISF  
 SVDLHDHPSIEDLTKSKYRIFENQTEDEFLINDIYVELKTPSRARKYLFSEIYN  
 SDSVYKDIRLEFNPKNLKIRGLHNVYASVSLIALTLGLKPEDFEVYIFPGILPR  
 LEFLNGNVEVYNDKSTTPHALMALKTPEPDSVILIVGKDKGADFEYSLRIYVK  
 KVIJALAGETKEKIDSKWDITEVKTONTLEAVKLAREVSLGNVYLFSPACSPD  
 MFRNVEERGEKEFLVEIWPRT"  
 complement(6382..6996)  
 /gene="trpP"  
 /note="aq\_2076"  
 complement(6382..6996)  
 /gene="trpP"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phosphoribosyl anthranilate isomerase"  
 /protein\_id="AAC07811.1"  
 /db\_xref="GI:2984287"  
 /translation="MWKVKICGTINLEDAFCAGADYIGVITPYKSPRYTSKRL  
 EIIKALELNVKVAVNVNPEYEFIKELDGFLLIDHGEDIELEKRGVLDKIVY  
 FRVKEIPKVEGEMKAVAILDFTSKRAYGSGTKTFMWEIAKLEVEGSHKVLISGLN  
 PENYKEGEIEFYPGVYSSGVEKEKKDKKVRREFIKTKSE"  
 7015..8556  
 /gene="snf"  
 /note="aq\_2077"  
 7015..8556  
 /gene="snf"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Na(+):neurotransmitter symporter (Snf family)"  
 /protein\_id="AAC07817.1"  
 /db\_xref="GI:2984293"  
 /translation="MEVKREHMAVRLGILMAGNAVGLNFLRPEYQAAENGGAEM  
 IPYIATFLVGLPIEMAMRGVAGAGSHGTTTPIAYELLRNRFPAKILGFGAMIP  
 VYAIYVYVESMTLGFATKPLVGLVPEPPNATDDSLRFKELVSIYIGVGGDEP  
 ILKPSLFAIYVFLITMFLINVELLIRGISKGIERRAKTAMPFLTLAVTVIRVLIET  
 PNGTADGULNLPDPEFKLDPGVIAVAGQIFFTLSLFGALITTAATYRKQDIY  
 LSGTLPAFLNEKAVIIGGSIIPAFAVFAVANAVALAKGAINLGLTLPALFSOT  
 AGGTGLGFLMEFLFAGLTSSIAIMQPIAFLEDELKLSKHAVLMTAAVLFPSAHL  
 VWEFLKSLDEMFNAGTIGVVFGLTELIFEFMLFGADKAMEINRGQIILKVPYIY  
 VMRYTTPAFVAVLAVMAREVYIPKIMETHTMTVITRYIIGLFLFLFVFLAERRR  
 NHESA"  
 8543..9394  
 /gene="sppA"  
 /note="aq\_2080"  
 8543..9394  
 /gene="sppA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="proteinase IV"  
 /protein\_id="AAC07815.1"  
 /db\_xref="GI:2984291"  
 /translation="MRVLKMLLIFSLILIFSVEKEFLSKFPVGHRIAVLKLGSVI  
 NPYLINKEIKRDKSNVAIVLRYVDSGGSGASQDIYRAIERKSKPVVISMGN  
 VASGGIYVSAAGNIIFANESTITSGIYIIOHVAKKLVEKLGIEATIAITGKFKDT

LSPRELTOEKEYLONLVNLEAQFIQAILIKYRSNKISEEKLREIADGRIFTGROAL  
 ELGIVDELGSUTDAIEKAKELSGYDAPVPFIOEERKSLKRLVADALEIKIINAYTP  
 LFLYLDY"  
 9672..11285  
 /gene="aq\_2082"  
 9672..11285  
 /gene="aq\_2082"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAC07820.1"  
 /db\_xref="GI:2984296"  
 /translation="MKRVLEFLGFGGLVGFVFVGVMSIIGLHLEHPPLYNAGF  
 ASVFLVLSFLNLINWVKALIKDYKAGSVYAFYDFLASIKLAIETMLVGLISML  
 GSTYIKONOSFEMYLDOFGYDIWIMKMLMDVHWSYIILFYLILAVNLIPCSIKR  
 LPRVMQAFSKREIRIILKDEHAKHKLPIYKIPDKVLKFLKGFVFEENKML

Query Match 99.8%; Score 1063.4; DB 1; Length 12497;  
 Best Local Similarity 99.9%; Pred. No. 5.1e-290;  
 Matches 1064; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	ATGATACCCAGAGATTAAAGACTTGAAGCTTAAAGACGAGGCTCACTCCGCTCC	60
Db	12335	ATGATACCCAGAGATTAAAGACTTGAAGCTTAAAGACGAGGCTCACTCCGCTCC	12276
QY	61	GTCAGGCTTCTCTAAGCAATTCCTACGACTTCCGAGAGATTAACAAGAGGCC	120
Db	12275	GTCAGGCTTCTCTAAGCAATTCCTACGACTTCCGAGAGATTAACAAGAGGCC	12216
QY	121	TTGAAGAATTAAAAAGTTCCTTGAACAATPACCCAGACCCGAGGAGAGTTA	180
Db	12215	TTGAAGAATTAAAAAGTTCCTTGAACAATPACCCAGACCCGAGGAGAGTTA	12156
QY	181	AAACGGCTTCTGCGGATTTTTCGCGGTTAAGAGAAAAATTAGTTCGCTAAGCT	240
Db	12155	AAACGGCTTCTGCGGATTTTTCGCGGTTAAGAGAAAAATTAGTTCGCTAAGCT	12096
QY	241	TGCGAGCACTCATATCTACCTCTCATAGCTATAGTGAACCTTACATACCGTTAC	300
Db	12095	TGCGAGCACTCATATCTACCTCTCATAGCTATAGTGAACCTTACATACCGTTAC	12036
QY	301	ATACCTGTCCACCTTCCCATGATAGATTAAGTTCGGAAGTTCGGAAGACCCCTC	360
Db	12035	ATACCTGTCCACCTTCCCATGATAGATTAAGTTCGGAAGTTCGGAAGACCCCTC	11976
QY	361	GTAAGGTTCAACTGACGAAAACTTGAATAGACTTGAAGAAGATTTGAATTAATA	420
Db	11975	GTAAGGTTCAACTGACGAAAACTTGAATAGACTTGAAGAAGATTTGAATTAATA	11916
QY	421	GAGAAAGAAAAACCCGTTCTCGGCTACTTGTACCCAAACAAACCCAGGAAACCTC	480
Db	11915	GAGAAAGAAAAACCCGTTCTCGGCTACTTGTACCCAAACAAACCCAGGAAACCTC	11856
QY	481	TTTTCACAGGGAAGATTGAAGATTAAGAAACAGSGGTCTTTCTGTATAAGAGAA	540
Db	11855	TTTTCACAGGGAAGATTGAAGATTAAGAAACAGSGGTCTTTCTGTATAAGAGAA	11796
QY	541	GCTACTATCACTTACTCGGAGAAACCTTCTGGAAGACGCGCTCAAAAGGAGATACG	600
Db	11795	GCTACTATCACTTACTCGGAGAAACCTTCTGGAAGACGCGCTCAAAAGGAGATACG	11736
QY	601	GTAGTTTGGAGACACTTCAAAAATCGGATAGCGGATTAAGGATGAGGATTTTATA	660
Db	11735	GTAGTTTGGAGACACTTCAAAAATCGGATAGCGGATTAAGGATGAGGATTTTATA	11676
QY	661	GGAAGGGGGAATCGTCTCAGAAATTACAGGTGAGACTCCCTTCAACGTGACCTAC	720
Db	11675	GGAAGGGGGAATCGTCTCAGAAATTACAGGTGAGACTCCCTTCAACGTGACCTAC	11616
QY	721	CCCTCTCAGGTGATGCAAAAGTTCTCTCAGGAGGAGAGAGATTCCTATAGGAAAG	780
Db	11615	CCCTCTCAGGTGATGCAAAAGTTCTCTCAGGAGGAGAGAGATTCCTATAGGAAAG	11556

QY	781	ATACAGGAGCTGTGTAAAGCGGAGAAAGATGTACACGCAAAATGAAAGAAATGAAAGGA	840		
Db	11555	ATACAGGAGCTGTGTAAAGCGGAGAAAGATGTACACGCAAAATGAAAGAAATGAAAGGA	11496		
QY	841	GTTGAGGTTTTCCAGTAAGGCTTAACCTTCTTGCTTTTCACACGCCCTTACCCGGCCAC	900		
Db	11495	TTTTGAGGTTTTCCAGTAAGGCTTAACCTTCTTGCTTTTCACACGCCCTTACCCGGCCAC	11436		
QY	901	GAGGTTTATCAGAGACTCTGTAAGGAGATGCTCTCGTACAGAAAGTATCTTACATGAA	960		
Db	11435	GAGGTTTATCAGAGACTCTGTAAGGAGATGCTCTCGTACAGAAAGTATCTTACATGAA	11376		
QY	961	GGACTGCCAAAGTGCTCAGGGGTAGCGTAGGAAACCGGAGAAACACACAGTTTCTG	1020		
Db	11375	GGACTGCCAAAGTGCTCAGGGGTAGCGTAGGAAACCGGAGAAACACACAGTTTCTG	11316		
QY	1021	GAGCAGCTGGAGAGAGATATAATCCCTTTCAAGCTCTCTTTAA	1065		
Db	11315	GAGCAGCTGGAGAGAGATATAATCCCTTTCAAGCTCTCTTTAA	11271		
RESUT 4	AE013160/c	12075 bp	DNA	linear	BCF 09-MAY-2002
LOCUS	DEFINITION	Thermoanaerobacter tengcongensis strain MB4T, section 187 of 244 of the complete genome.			
ACCESSION	VERSION	AE013160	AE008691		
KEYWORDS	AE013160.1	GI:20517144			
SOURCE	ORGANISM	Thermoanaerobacter tengcongensis.			
REFERENCE	AUTHORS	Thermoanaerobacter tengcongensis Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.			
1 (bases 1 to 12075)					
QY	1021	Bao, Q., Tian, Y., Li, W., Xu, Z., Xuan, Z., Hu, S., Dong, W., Yang, J., Chen, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y., Ling, L., Tan, H., Chen, R., Wang, J., Yu, J., and Yang, H.			
QY	1021	A Complete Sequence of the T. tengcongensis Genome			
QY	1021	Genome Res. 12 (5), 689-700 (2002)			
QY	1021	21992816			
QY	1021	11997336			
QY	1021	2 (bases 1 to 12075)			
QY	1021	Bao, Q., Xu, Z., Hu, S., Dong, W., Chen, Y., Wang, J., Yu, J., and Yang, H.			
QY	1021	Direct Submission			
QY	1021	Submitted (07-OCT-2001) Beijing Genomics Institute/Genomics and Bioinformatics Center, Institute of Airports and Development, Chinese Academy of Sciences, Beijing Airport Industrial Zone B6, Beijing 101300, China			
QY	1021	3 (bases 1 to 12075)			
QY	1021	Li, W., Xuan, Z., Yang, J., Ling, L., and Chen, R.			
QY	1021	Direct Submission			
QY	1021	Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China			
QY	1021	4 (bases 1 to 12075)			
QY	1021	Tian, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y., and Tan, H.			
QY	1021	Direct Submission			
QY	1021	Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, China			
QY	1021	Location/Qualifiers			
QY	1021	1. 12075			
QY	1021	/organism="Thermoanaerobacter tengcongensis"			
QY	1021	/strain="MB4T"			
QY	1021	/db_xref="taxon:119072"			
QY	1021	115_143			
QY	1021	/note="putative rho-independent transcription terminator"			
QY	1021	complement(145..781)			
QY	1021	/gene="HisI"			
QY	1021	/note="TTE2132"			
QY	1021	complement(145..762)			
QY	1021	/gene="HisI"			
QY	1021	/note="Best Blastp hit =			
QY	1021	9416175609[gb]AAAF05093.1[AF150930.4 (AF150930)			
QY	1021	phosphoribosyl-AMP cyclohydrolase [Thermoanaerobacter			
QY	1021	ethanolicus], score 237, E-value 8.00E-62"			
QY	1021	/codon_start=1			
QY	1021	/transl_table=11			
QY	1021	/product="Phosphoribosyl-AMP cyclohydrolase"			
QY	1021	/protein_id="AA025297.1"			
QY	1021	/db_xref="GI:20517145"			
QY	1021	/translation="MIDIKELKFDKGLIPAYVDYPTKQVLMALYMNESKLTLE			
QY	1021	GEYFFSRSGKIMHGKSGNTQKVKLFYQCDDEALLIOVEAGPACGHNISCFE			
QY	1021	RSPEETDEGIELINKYERIKGRKINPEGSGYVLPKGIKTLKVGAEAEVYI			
QY	1021	ASKNDSKDEIVYVSDLIYHMLVLEKNTITLNDIYNLEKTKK"			
QY	1021	complement(151..417)			
QY	1021	/gene="HisI"			
QY	1021	/note="Pfam match to entry PRA-PH, Phosphoribosyl-ATP			
QY	1021	pyrophosphohydrolase, score 159.1, E-value 7.50E-44"			
QY	1021	complement(445..678)			
QY	1021	/gene="HisI"			
QY	1021	/note="Pfam match to entry PRA-CH, Phosphoribosyl-AMP			
QY	1021	cyclohydrolase, score 186.6, E-value 4.10E-52"			
QY	1021	complement(777..781)			
QY	1021	/gene="HisI"			
QY	1021	complement(780..1557)			
QY	1021	/gene="HisF"			
QY	1021	/note="TTE2133"			
QY	1021	complement(780..1541)			
QY	1021	/gene="HisF"			
QY	1021	/note="Best Blastp hit =			
QY	1021				

```
CDS      complement(2255. .2866)
/ gene="Hish"
/ note="Best Blastp hit = g112913863|sp|O66943|HIS5_AQUAE
AMIDOTRANSFERASE HISH g117433986|pir|C70364
amidotransferase Hish - Aquifex aeolicus
g112983311|gb|AAC06901.1| (AE000771) amidotransferase Hish
[Aquifex aeolicus], score 207, E-value 7.00E-53"
/codon_start=1
/transl_table=11
/product="Glutamine amidotransferase"
/protein_id="AAM25300.1"
/db_xref="GI:20517149"
/translation="MIGIDYGMGNLRSVERAKLYGYDAKITSDAKEIKNFALLP
GVGAPPAIRVLITGMAREIKIHIEKGFELGICMOLLFEKSFEGKTEGKIFK
GNVGLPKGNKIPHYGNASLTIKKDTPGICIKGDYVPAHYSVYNGNKNDEYAVT
EYGIIPAVIITKDVVSCQFPEKSGETGLKILNFGEMIKCL"
complement(2267. .2860)
/ gene="Hish"
/ note="Pfam match to entry GATase, Glutamine
amidotransferase class-I, score 176.7, E-value 3.80E-49"
complement(2275. .2279)
/ gene="Hisa"
complement(2874. .2878)
/ gene="Hish"
complement(2885. .3482)
/ gene="Hish"
/ note="TTE2136"
complement(2885. .3469)
/ gene="Hish"
/ note="Best Blastp hit = g117387773|sp|O33564|HIS7_RHOSH
IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE (IGDP)
g112398791|emb|CA60712.1| (X87256)
imidazoleglycerol-phosphate dehydratase [Rhodobacter
sphaeroides], score 223, E-value 2.00E-57"
/codon_start=1
/transl_table=11
/product="imidazoleglycerol-phosphate dehydratase"
/protein_id="AAM25301.1"
/db_xref="GI:20517149"
/translation="MRKAVERKRTAETDIYELNIDKGSYDIKIGFEDMLSLPA
KHGFLDKVIAKGDEVDHNVGIVLGATKAGADKSKIRFETFPVPEAL
VRVSLDISGRPLIYDPLKARVGNFETFEVFFRFAVFGITLHVELLHGVNH
HIEASFALGADDEALKDERIDIGISTGIL"
complement(2945. .3379)
/ gene="Hish"
/ note="Pfam match to entry IGDP,
imidazoleglycerol-phosphate dehydratase, score 293.4,
E-value 6.20E-100"
complement(3466. .4534)
/ gene="Hisc2"
/ note="TTE2137"
complement(3466. .4521)
/ gene="Hisc2"
/ note="Best Blastp hit = g117450424|pir|E70478
hisdinol-phosphate aminotransferase - Aquifex aeolicus
g112984289|gb|AAC07813.1| (AE000771) hisdinol-phosphate
aminotransferase [Aquifex aeolicus], score 230, E-value
2.00E-59"
/codon_start=1
/transl_table=11
/product="Hisdinol-phosphate aminotransferase/Tyrosine
aminotransferase"
/protein_id="AAM25302.1"
/db_xref="GI:20517150"
/translation="MIENLREIKGFKNVEVNPYKKNMANTPEELPEEVKNI
GTVSIHNITPDPAETKREIARVCSVPPKNI FVGNSDEITHIIMLAFVKGDT
VLYPHEFAMSIYISKIAGNAIYAVLNEDYTYNPREAVERKRPKLVPCNPNP
TGSVIDEEDIRITKARGIVIVDAVFEFGTLVPYIDRENLILRTSKAGTIA
GLRVGALNNGEIVKYLNVKSPYNLNSQRIAEVLKSVGLKERVNTIINERKLV
KELNGLNGIKVYPSHANEVLCFENANDVHKRLVRGLVRNFSNVKGLGTLRTIVS
SSDANDYLINALREILS"
complement(3469. .4515)
/ gene="Hisc2"

misc_feature
complement(2255. .2866)
/ gene="Hish"
/ note="Best Blastp hit = g112913863|sp|O66943|HIS5_AQUAE
AMIDOTRANSFERASE HISH g117433986|pir|C70364
amidotransferase Hish - Aquifex aeolicus
g112983311|gb|AAC06901.1| (AE000771) amidotransferase Hish
[Aquifex aeolicus], score 207, E-value 7.00E-53"
/codon_start=1
/transl_table=11
/product="Glutamine amidotransferase"
/protein_id="AAM25300.1"
/db_xref="GI:20517149"
/translation="MIGIDYGMGNLRSVERAKLYGYDAKITSDAKEIKNFALLP
GVGAPPAIRVLITGMAREIKIHIEKGFELGICMOLLFEKSFEGKTEGKIFK
GNVGLPKGNKIPHYGNASLTIKKDTPGICIKGDYVPAHYSVYNGNKNDEYAVT
EYGIIPAVIITKDVVSCQFPEKSGETGLKILNFGEMIKCL"
complement(2267. .2860)
/ gene="Hish"
/ note="Pfam match to entry GATase, Glutamine
amidotransferase class-I, score 176.7, E-value 3.80E-49"
complement(2275. .2279)
/ gene="Hisa"
complement(2874. .2878)
/ gene="Hish"
complement(2885. .3482)
/ gene="Hish"
/ note="TTE2136"
complement(2885. .3469)
/ gene="Hish"
/ note="Best Blastp hit = g117387773|sp|O33564|HIS7_RHOSH
IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE (IGDP)
g112398791|emb|CA60712.1| (X87256)
imidazoleglycerol-phosphate dehydratase [Rhodobacter
sphaeroides], score 223, E-value 2.00E-57"
/codon_start=1
/transl_table=11
/product="imidazoleglycerol-phosphate dehydratase"
/protein_id="AAM25301.1"
/db_xref="GI:20517149"
/translation="MRKAVERKRTAETDIYELNIDKGSYDIKIGFEDMLSLPA
KHGFLDKVIAKGDEVDHNVGIVLGATKAGADKSKIRFETFPVPEAL
VRVSLDISGRPLIYDPLKARVGNFETFEVFFRFAVFGITLHVELLHGVNH
HIEASFALGADDEALKDERIDIGISTGIL"
complement(2945. .3379)
/ gene="Hish"
/ note="Pfam match to entry IGDP,
imidazoleglycerol-phosphate dehydratase, score 293.4,
E-value 6.20E-100"
complement(3466. .4534)
/ gene="Hisc2"
/ note="TTE2137"
complement(3466. .4521)
/ gene="Hisc2"
/ note="Best Blastp hit = g117450424|pir|E70478
hisdinol-phosphate aminotransferase - Aquifex aeolicus
g112984289|gb|AAC07813.1| (AE000771) hisdinol-phosphate
aminotransferase [Aquifex aeolicus], score 230, E-value
2.00E-59"
/codon_start=1
/transl_table=11
/product="Hisdinol-phosphate aminotransferase/Tyrosine
aminotransferase"
/protein_id="AAM25302.1"
/db_xref="GI:20517150"
/translation="MIENLREIKGFKNVEVNPYKKNMANTPEELPEEVKNI
GTVSIHNITPDPAETKREIARVCSVPPKNI FVGNSDEITHIIMLAFVKGDT
VLYPHEFAMSIYISKIAGNAIYAVLNEDYTYNPREAVERKRPKLVPCNPNP
TGSVIDEEDIRITKARGIVIVDAVFEFGTLVPYIDRENLILRTSKAGTIA
GLRVGALNNGEIVKYLNVKSPYNLNSQRIAEVLKSVGLKERVNTIINERKLV
KELNGLNGIKVYPSHANEVLCFENANDVHKRLVRGLVRNFSNVKGLGTLRTIVS
SSDANDYLINALREILS"
complement(3469. .4515)
/ gene="Hisc2"

RBS
/ note="Pfam match to entry aminotran_1, Aminotransferase
class-I, score -25.9, E-value 1.60E-08"
complement(3478. .3482)
/ gene="Hish"

Query Match      13.7%: Score 145.4: DB 1: Length 12075:
Best local similarity 49.8%: Pred. No. 7.5e-30:
Matches 511: Conservative 0: Mismatches 501: Indels 15: Gaps 5:

16  ATTAGAAGACTTGAAGCTTACAGAGGAGGTCACGCCCCCGCTCGAGGCTTCTCT 75
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4494  ATTTAAAGAGTTTAAATAATTTATGAATGGAATAATGCTTTATTAATTAAGATGGA 4435
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
76  AACCAATTCCTCCACGACTTTCGAGAGATTAACCAAGGCGTTAGAAGATTAA 135
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4434  AAGCAGACCCCTTTTGAACCTCCTGAAGAGATGAATAAATATAGGGATATTGAAG 4375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
136  AAGTTCCTCCCTGAACCAATATCCAGACCCCGAAGCCGAAGAGTTAAAGCGTTTCGG 195
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4374  TCTATCCATGTAAATATATATATCCAGACCCCTGAGCAAAAAATTTGAGAGGTTAGCG 4315
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
196  GATTTTTCGCGCTTAAGAGAAATTTAGTTCGCTGCTGCTGCGAGCACTCAT 255
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4314  AGATATTGCACTGCTCACTCCTAAATAATTTTGTAGAAATGTTCTGATGAATAATA 4255
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
256  TACTACCTTCATATAGCTATAGTGAACCTTACATACCGCTTTACATCTGTTCCAC 315
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4254  CACCTTATATAGCTTGGCTTGTGATAAAGGAGATACAGTTTGTGACCTCCACATCT 4195
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
316  TTTCCCATGTACGATATAGTCCGAAGTCTCTGAGAGACCCCTCGTAAGTTCACTG 375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4194  TTTCCATGTACGATATAGTCCGAAGTCTCTGAGAGACCCCTCGTAAGTTCACTG 4135
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
376  GACCAAACTTGTATATAGCTTATAGAAAGATATGAATTAATAGAGAAAGAAACC 435
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4134  AATGAGACATTAATTAATATGTTGAAGATTTCCAGACAGATGAAGATTAAGGCT 4075
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
436  GTTCTCGGACTTGTGCTTACCCCAACACCCCGGAAACCTCTT--TCAGGGGA 493
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4074  AAATTTGGTTTGTGCAATCCCAACACATCCCTGAGAGCGTAAATAGTAGAGAGAT 4015
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
494  AGATTGAGAGATTAAGAACAGG-GTGTCTTGTGTAAATAGACGACTATCAT 552
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4014  ATATACAGATTAATTTGAAGAAAGCAAGAGGATTTATATAGTATGAAGCTTATTGAA 3955
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
553  TACTCCGAGAACCTTTCGAGAGCGGCTCAAAAGGAGATACGGATTTTGAGG 612
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3954  TTTATGAGAAACCTTTGTGCGCATATATAGATTTGAAATTTGATTTGCTGCGA 3895
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
613  ACACCTTCAGAA--ATGCGTATGCGAGTTTAAAGGATTTTAAAGGAAGGG 669
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3894  ACATTGTCAAAAGCTTTGGGATTGCAAGATTAAGAGTGGGATATGCTCTCAATGCA 3835
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
670  GAAATCGTCAAGAAATTAACAGAGTGAGACTCCCTTCAACGTGACCTACCCCTCAG 729
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3834  GAAATAGGAATATTTTAATCTGTAAATCCTTATATCACTCTTATACAG 3775
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
730  GTGATGCAAAAGTTCTCCTCAGGAGGAGAGAAATTCCTAATGGAAGAAAGTACAGAG 789
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3774  AGAATAGCATTTAGAAAGTTTAA-----AAGTGTGTTTAAAGGAGAGGTTAAATTAC 3721
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
790  GTTGAACAGAGCCGAGAAAGATGACAGAAATGAAGAAATAGAGAGATGAGAGTT 849
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3730  ATTTAAAGCAAGAGAAAGCTGTTAAGAGATTAATTAATTAATTAATGATGATAAGTG 3661
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
850  TTTCCAGTAAAGCTTAACCTTTCGTTTGAACCGCTTACCCGCCACAGAGTTTAT 909
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3660  TATCTTCTCAAGCAATTTTGTCTT--GTAATTTGAAGAAATGCAATGATGTCAC 3604
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
910  CAGGACTACTGAAAGAGATGCTCTCTCAGAGACGATCTTACATGGAAGAGACTCAA 969
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3603  AAAAGACTTGTGAAAGGGAATTTTGGTGAAGAAATTTTCTAATGTGAAGGGGTTGAA 3544
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```



Histidine family: (hisg)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Atp phosphoribosyltransferase"  
 /protein\_id="AAL81782.1"  
 /db\_xref="GI:18893815"  
 /translation="MREYVLPKGRLYQDSIKYVELAGIKVPRKDERALIMGDSNNFELL  
 AAFDPVATIKTIGIAGSDVNEBVSDFPIPLEPGKRLSIAPKREYVDPEN  
 MCGKATATYVNTIKRYFENLNIEVEVRIKLHSGIELAPRIGIADIVIVETGNTLRL  
 NGLVEAKIMDVYALLVNRISQKMFEEINSEITKIRSYVGH"  
 1614. .2741  
 /gene="PF1659"  
 1614. .2741  
 /gene="PF1659"  
 /EC\_number="3.5.4.19"  
 /note="Function Code: 1.7 Amino Acid Biosynthesis:  
 Histidine family"  
 /codon\_start=1  
 /transl\_table=11  
 /product="histidinol dehydrogenase"  
 /protein\_id="AAL81783.1"  
 /db\_xref="GI:18893816"  
 /translation="MDIEYVRELLKDIKREGLAEIREYSKKFDCEEPFEYSKEEF  
 EKEERIPEDKRTIKRTLKREIKYRKQDKDITLENGSLYGLIFRPLKRIQIVPP  
 GPPLSLTMVGVPARVAVGKEIVITNPDKQKVPYLLIYAKLRIEYVRLGIGIKGIVP  
 IAAWYGLIMGKRVQKIFGCGNFVNEAKROYGVPIGIDSLGSPSEIAVIAESADKDD  
 VYLDLSQLEHSGDSKRAKMLITTSKEIADYCNNEGITYVCSLEECCKANIEAHEH  
 EITTRPELVDLLENAGAILYGLTPADYVFGVNHVLPYTGAAKFMGVLTIVWDE  
 MPDISIAWSEREFPISERELTGRILAEIGIELHKSKLEVR"  
 2808. .3338  
 /gene="PF1660"  
 2808. .3338  
 /gene="PF1660"  
 /EC\_number="4.2.1.19"  
 /note="Function Code: 1.7 Amino Acid Biosynthesis:  
 Histidine family"  
 /codon\_start=1  
 /transl\_table=11  
 /product="imidazoleglycerol-phosphate dehydratase"  
 /protein\_id="AAL81784.1"  
 /db\_xref="GI:18893817"  
 /translation="MRRTTKENDILVEIGKKGGEIKTNDLILDHMLTAFAPLIGDMR  
 TATYDLRHMEIDIGITIGEALRENLPRKPTRREGAIMPMDALIVIVSDISNPYAN  
 VDVMNLKADPEGRVAVSLLEFVWGLARGLRATITHIQSLGSENAHHIVEAEFGILGMALE  
 VATKSERVESTKGVL"  
 3338. .3922  
 /gene="PF1661"  
 3338. .3922  
 /gene="PF1661"  
 /note="Function Code: 1.7 Amino Acid Biosynthesis:  
 Histidine family (hisH)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="glutamine amidotransferase"  
 /protein\_id="AAL81785.1"  
 /db\_xref="GI:18893818"  
 /translation="MDRIAYVDIGLGNLANYKKAAGYITSDPYEIEKADYIVPGV  
 NGAAYDKLAPRKDILITIGESINBGRPFQICIGMOLLFESESESPKBSLGIFKRVVAK  
 LKNVTRPHGMQVNIKKCKLFESLKNGSYFVYHSAVHAPDDIILATTTDENME  
 EVSSYSCFENEGVDPHPEKSKSNGDILIRNFRRL"  
 3924. .4646  
 /gene="PF1662"  
 3924. .4646  
 /gene="PF1662"  
 /EC\_number="2.3.1.22"  
 /note="Function Code: 1.7 Amino Acid Biosynthesis:  
 Histidine family"  
 /codon\_start=1  
 /transl\_table=11  
 /product="n-(5'-phospho-d-ribosyl)formimino--5-amino-1-  
 5'-phosphoribosyl)-4-imidazolecarboxamide isomerase"  
 /protein\_id="AAL81786.1"  
 /db\_xref="GI:18893819"  
 /translation="MKFRITYPADIMDGKVVVRLYGGKKDEYKVGDPPLKFAEKSESEYV





NRFEHFWASRSDPVSGETIDRKRKIAISCKKLGRIYIISYLHEVEALAAASALIH  
 FKSVLGKKLKNKTAGHSPGISTVEIDRREGESLPLNEINPGLAIKINAPDL  
 DTDELDLYLENDODANLKSRTLSNWTGKDVASKVYIGVGTGIONAVYSAYEDG  
 FEIWEETKNGRKOSRYRERTTSSEEVLAQLGQKCLPELKFSGSGEAYLRKPE  
 TETQENIQTTPNPKIKKEKTKS"  
 8509. 9798  
 /gene="NM1581"  
 8509. 9798  
 /note="similar to GB:AE000657 percent identity: 71.03;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="histidinol dehydrogenase"  
 /protein\_id="AA041934.1"  
 /db\_xref="GI:726829"

translation="MKKLINTSPDPOGLKALLAFETAQNTERIVADICADYOKRG  
 DAALLEYNKEPDQINAKSIDLILTOADLNAAFERINDVOTLALOTARVESHQ  
 KMESSYDEDEGLIGQIPELDKRVGIVPCGKAAYSSVIMNMPAHVAGVEIIV  
 VTPKGERNDIVIAAAYVAGVKTETVGAQVAVALAYGETTIPQVDRKINGPNAFVA  
 AAKRVEGVGIDWVAGPSELIVLADGTPADWVAMDLFSAEHDELAQALITGSA

Query Match 8.5%; Score 90.8; DB 1; Length 11932;  
 Best Local Similarity 47.7%; Pred. No. 2.2e-14;  
 Matches 331; Conservative 0; Mismatches 357; Indels 6; Gaps 2;

QY 182 AAGCGTTCTCGGATTTTTCGGGCTAAGAGAAATTTAGTTCGGTACGGT 241  
 DB 10058 AAGCATTTAGTTCGGGCTTGCATATCCCGACGACATTCGGCTGGCAACGGTT 10117  
 QY 242 CGGACGACATCATATCTACTCTCATATAGTATAGTTCATATACATACCGCTTTACA 301  
 DB 10118 CGGACGACATCATATCTACTCTCATATAGTATAGTTCATATACATACCGCTTTACA 301  
 QY 302 TACCTTTCCGACCTTCCCATGTAGAGATATAGTTCGGAATTTCTCGGAAGACCCCTCG 361  
 DB 10178 CAGCCGACACCACTTTCGATCATACGCGCACACGCGCGCTGATGCGCATATATG 10237  
 QY 362 TAAAGTTCACTGACGAGAAATCTTGATATAGTTCATAGAAAGATTTAGATTAAATG 421  
 DB 10238 TCGGCGTTCCATCAAGGAGATTTACCTCACTGCGCCGCTGCTCGAAGCGGTCA 10297  
 QY 422 AGAAGAAAAACCGCTTCGCGGTTACTTGTACCCAAACACCCGACGGGAAACCTCT 481  
 DB 10298 GGAACACCGCCCTGCGCTGACTTATGCGCTACCCCAACACCCGACGGGATGCT 10357  
 QY 482 TTTCCAGGGGAAAGATTGA---GGAGATAGAAACAGGGGTTTCTGTGTAATAGAG 538  
 DB 10358 TCACGGGTGCGGAATCGAAGCGGTGATCGAAGCTTCAGACGGCATCTGTCTGATG 10417  
 QY 539 AAGCTTACTATCATTTACTCGCGAGAAACCTTCTGGAAGACGGGTCAAAAGGAGATA 598  
 DB 10418 AAGCTTACGGCGATTCACAGCGGACAGCTTCTGCGCGACGAGGACAGATTTCCCAAC 10477  
 QY 599 CGGTAGTTTGGAGACATTTCAAAATCGTATGCGGATTTAAGGTTAGGATTTTAA 658  
 DB 10478 TGAATGCTTACGACCCCTGACGAAATCGGTTTTCGCGACTGCTATCGGTTATGCGG 10537  
 QY 659 TAGGAGAGGGGAAATCGTCTGAAATTTAAGAGTGAAGTCCCTTCACAGTACCT 718  
 DB 10538 CAGGCTGCGCCGGAAGTATGCGGAAATTCCTGCGCCCTTACATATATACAC 10597  
 QY 719 ACCCTCTCAGTGTATGCGCAAAAGTTCTCTACGAGGAGGAGAGATTTCTTAATGAAA 778  
 DB 10598 AATTGAGCTGACCACTGCGCAAACTCGCTGGGCACTACGCGCATATCTCT---GCCA 10654  
 QY 779 AGATACAGAGGTTGTAAACAGAGCGAAGATGTACGAGAAATGAAACAAATAGAG 838  
 DB 10655 ACATCGACAGCTGAAAGAGAGAGCGATGTTTCGCGAATTTGGCAAAATATATGCC 10714  
 QY 839 GAGTTGAGTTTTCGAGATAGGCTACTTCTT 872  
 DB 10715 GTCTGAACACCTTTTCAAGTCAGGCAAACTTCAT 10748

RESULT 7  
 AX044033 349980 bp DNA linear PAT 24-NOV-2000  
 LOCUS  
 DEFINITION  
 AX044033  
 AX044033  
 AX044033.1 GI:11342917  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Neisseria meningitidis.  
 Neisseria meningitidis  
 Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;  
 Neisseria.  
 1 (bases 1 to 349980)  
 Pizzaz, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J.C.,  
 Masignani, V., Galeotti, C., Mora, M., Ratti, G., Scarselli, M.,  
 Scarlato, V., Rappuoli, R., Frazer, C.M. and Grandi, G.  
 Neisseria genomic sequences and methods of their use  
 Patent: WO 006791-A 112 09-NOV-2000;  
 CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)  
 Location/Qualifiers  
 1. 349980  
 /organism="Neisseria meningitidis"  
 /db\_xref="taxon:487"  
 /note="sequence too long, cut in 8 pieces, -seq 1: 1 to  
 349980 349980 bases -seq 108: 300001 to 649980 349980  
 bases -seq 109: 600001 to 949980 349980 bases -seq 110:  
 900001 to 1249980 349980 bases -seq 111: 1200001 to 1549980  
 349980 bases -seq 112: 1500001 to 1849980 349980 bases -seq  
 113: 1800001 to 2149980 349980 bases -seq 114: 2100001 to  
 2272325 172325 bases"

BASE COUNT 87189 a 93501 c 84627 g 84663 t  
 ORIGIN

Query Match 8.5%; Score 90.8; DB 6; Length 349980;  
 Best Local Similarity 47.7%; Pred. No. 3.1e-14;  
 Matches 331; Conservative 0; Mismatches 357; Indels 6; Gaps 2;

QY 182 AAGCGTTCTCGGATTTTTCGGGCTTAAAGAAATTTAGTTCGGTACGGT 241  
 DB 144619 AAGCATTTAGTTCGGGCTTGCATATCCCGACGACATTCGGCTGGCAACGGTT 144678  
 QY 242 CGGACGACATCATATCTACTCTCATATAGTTCATAGTTCATATACATACCGCTTTACA 301  
 DB 14479 CAGGACGACATCATATCTACTCTCATATAGTTCATAGTTCATATACATACCGCTTTACA 301  
 QY 302 TACCTTTCCGACCTTCCCATGTAGAGATATAGTTCGGAATTTCTCGGAAGACCCCTCG 361  
 DB 144739 CAGCCGACACCACTTTCGATCATACGCGCACACGCGCGCTGATGCGCATATATG 144798  
 QY 362 TAAAGTTCACTGACGAGAAATCTTGATATAGTTCATAGAAAGATTTAGATTAAATG 421  
 DB 144799 TCGGCGTTCCATCAAGGAGATTTACCTCACTGCGCCGCTGCTCGAAGCGGTCA 144858  
 QY 422 AGAAGAAAAACCGCTTCGCGGTTACTTGTACCCAAACACCCGACGGGAAACCTCT 481  
 DB 144859 GGAACACCGCCCTGCGCTGACTTATGCGCTACCCCAACACCCGACGGGATGCT 144918  
 QY 482 TTTCCAGGGGAAAGATTGA---GGAGATAGAAACAGGGGTTTCTGTGTAATAGAG 538  
 DB 144919 TCACGGGTGCGGAATCGAAGCGGTGATCGAAGCTTCAGACGGCATCTGTCTGATG 144978  
 QY 539 AAGCTTACTATCATTTACTCGCGAGAAACCTTCTGGAAGACGGGTCAAAAGGAGATA 598  
 DB 144979 AAGCTTACGGCGATTCACAGCGGACAGCTTCTGCGCGACTGCTATCGGTTATGCGG 145038  
 QY 599 CGGTAGTTTGGAGACATTTCAAAATCGTATGCGGATTTAAGGTTAGGATTTTAA 658  
 DB 145039 TGAATGCTTACGACCCCTGACGAAATCGGTTTTCGCGACTGCTATCGGTTATGCGG 145098  
 QY 659 TAGGAGAGGGGAAATCGTCTGAAATTTAAGAGTGAAGTCCCTTCACAGTACCT 718  
 DB 145099 CAGGCTGCGCCGGAAGTATGCGGAAATTCCTGCGCCCTTACATATATAGAAC 145158



gene  
GASLAPWKVA"  
5309..5644  
/gene="TM1027"  
5309..5644  
/gene="TM1027"  
/note="similar to percent identity: 0.00; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="A036104.1"  
/db\_xref="GI:4981568"  
/translation="MSSEPLGVGNIGARYSATYSMGTVPKAKELIGPSPGFIHP  
DLKDVLEIPIDIGFYWQNEPQSYGVSVITIFNGEPCIRACGYENGAAAEIKTYOTG  
LVIVQQLVE"  
5922..6803  
/gene="TM1028"  
5922..6803  
/gene="TM1028"  
/note="similar to GB:AE000782 percent identity: 60.48;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter, ATP-binding protein"  
/protein\_id="A036105.1"  
/db\_xref="GI:4981569"  
/translation="MSVVEYRDLTKYVGSRGVGYTFSEGEIFGPIGNGAKTT  
TIRLLGLIPDSGLAEITFGKDYLKKEKTRKRVGIIPEGVSPPVTVPEELRYSAS  
FYERVDVYKELCGVSLDPTKRRIKELSGNKKRYAIVQALHRRKLLIDEPITGL  
DPIVONTFPEILNKEKGTTFVSSHLISEVRLDVRAMIKDGI1IREKVENLNG  
ERYKVVVKGENLEKLDKLEAVNRLKFNENGTAFLEPSGNERLLLEIRLSKLSDFWE  
EPSEIEFMSYREGE"  
6800..7591  
/gene="TM1029"  
6800..7591  
/gene="TM1029"  
/note="similar to SP:P42333 percent identity: 50.27;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter, permease protein, putative"  
/protein\_id="A036106.1"  
/db\_xref="GI:4981570"  
/translation="MNIFRMDMKRYKSTIAWTVLLILQIMTAAPSPNAKTEFIT  
KMKVMPKAFKLEGLDMDSNIMNYLAWSSIVYLVGVGASIVGVSVEKEEKE  
KTVKELSKPVSREVIVKSFESLVHLLFDALFSLMFANITSSSVVEKEEYV  
PASEQVHLHTLNMISFPGVGTIRSDGALSTIGIVPVLVLMISKITDKAEFLKFT  
PESYADPSTVIRHGFPEFSGLEFALVINALILISSIVLFSKKDILV"  
7602..8204  
/gene="TM1030"  
7602..8204  
/gene="TM1030"  
/note="similar to GB:S87512 SP:P43506 PID:623036  
PID:247052 percent identity: 48.00; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator, TetR family"  
/protein\_id="A036107.1"  
/db\_xref="GI:4981571"  
/translation="WLSKRDAILKAAYVEFKKGYDRATVDEIAKAGVAKGLIFHF  
KNKEELYQAYMSVTEKLOKEFENFLMKNRRDIDFERMERTEKLEYSSHPPEAF  
LITLVSVDEGRKRIILDLKESORVEFDVREKLDLADVETETALAKFLWFFSG  
FEEVYLRTYQGPPELKRDMNTLVEEVYVLMRIIKGKMTK"  
complement(8229..8471)  
/gene="TM1031"  
complement(8229..8471)  
/gene="TM1031"  
/note="similar to GB:M60116 SP:P23171 PID:144907 percent  
identity: 73.61; identified by sequence similarity;  
Query Match 8.4%; Score 89.2; DB 1; Length 20334;  
Best Local Similarity 54.3%; Pred. No. 6.7e-14;

Matches 225; Conservative 0; Mismatches 183; Indels 6; Gaps 2;  
QY 427 GAAAACCCGTTCTCGGTACTTCTTACCCCAACACCCAGGGAACTCTTCC 486  
DB 15718 GGAGAGGAGACGTTGTTTCAATCCGAACCCGACATCCACAGGGCCATGTTGGA 15659  
QY 487 AGCGAAGATGAGAGATAAGAAACAGGGGTTTCTGTGTAATAGACAGCTAC 546  
DB 15658 ACAGAGGAATGAAAGATCTCGAAGACGGGTGCTGTCGGCTGAGAACCTAC 15599  
QY 547 TATCATTTACTCCGAGAAACCTTCTGAGACGGCTCAAAAGGAATACGTTAGTT 606  
DB 15598 TACGAATTCACAGAGAAAGTATGTGATTTCTGAAGAAATACGAAATCTGCTGTG 15539  
QY 607 TTGAGACACTTTCAAAA--ATCGGTATGCGAGCTTTAAGGTTAAGGATTTAATAGG 663  
DB 15538 ATCAGACACTTCTCGAAGACGTTTCCCGTAGGCAACGTTGGATGTTGTGCGC 15479  
QY 664 AAGGGGAATGCTCTCAGAAATTACAGATGACACTCCTTGAACGTACCTACCC 723  
DB 15478 TCGGAGAACTTATGACCTTACCAACAGGAGAGACTTCTTCAACGTGAGTACGTC 15419  
QY 724 TCTCAGCTGATGGCAAAAGTTCTCCACAGGAGGAGAGAAATTCATGCAAGAAATGATA 783  
DB 15418 TCCAGATGTTCCAGAGTAGTACTCTC---GATCACAGAGATCTTGAAGAAAGAACG 15362  
QY 784 CAGAGGTTGTACAGACGAGAGAGATGTACGAGAAATGACGAAATGACAA 837  
DB 15361 AATTTCATCGTGAAGACGCGAGAGATGAAGAGTCTCTCAGGAAATGGA 15308  
RESULT 9  
NMA622491 326301 bp DNA linear BCT 04-DEC-2000  
LOCUS  
DEFINITION  
Neisseria meningitidis serogroup A strain Z2491 complete genome;  
segment 6/7.  
ACCESSION  
AL162757.2 AL157959  
VERSION  
AL162757.2 GI:7380371  
KEYWORDS  
Neisseria meningitidis Z2491.  
SOURCE  
Neisseria meningitidis Z2491.  
ORGANISM  
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;  
Neisseria.  
REFERENCE  
1 (bases 1 to 326301)  
Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,  
Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,  
Davies,R.M., Davis,P., Devlin,K., Feltham,D., Hamlin,N.,  
Holliday,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,  
Quail,M.A., Raftery,M.A., Rutherford,K.M., Simmonds,M.,  
Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.  
Complete DNA sequence of a serogroup A strain of Neisseria  
meningitidis Z2491  
JOURNAL  
Nature 404 (6777), 502-506 (2000)  
MEDLINE  
20222556  
PUBMED  
10761919  
REFERENCE  
2 (bases 1 to 326301)  
Parkhill,J.  
TITLE  
Direct Submission  
AUTHORS  
Submitted on behalf of the Neisseria  
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA E-mail: parkhills@sanger.ac.uk  
COMMENT  
Notes:  
Details of N. meningitidis sequencing at the Sanger Centre are  
available on the World Wide Web  
(URL: [http://www.sanger.ac.uk/Projects/N\\_meningitidis/](http://www.sanger.ac.uk/Projects/N_meningitidis/)).  
FEATURES  
source  
1..326301  
/organism="Neisseria meningitidis Z2491"  
/strain="Z2491"  
/db\_xref="taxon:122587"  
/note="serogroup: A"  
196..311  
/note="REP 2; hmufs hit to HMM REP 2 (1 - 109), score:

repeat\_unit

```

repeat_unit      87.43"
                  434. .453
/note=">= 90% match to ATCCCNNNNNNGGAAAT"
repeat_unit      /label=dRS3
                  498. .560
/note="Correia element; hmms hit to HMM Correia (1 - 62),
                  score: 78.71"
                  /label=Correia
                  562. .604
/note="Correia element; hmms hit to HMM Correia (114 -
                  156), score: 51.92"
misc_feature      /label=Correia
                  complement(667. .676)
/note="Core DNA uptake sequence: gccgtctgaa"
gene              /label=DUS
                  complement(668. .1813)
                  /gene="dape"
                  complement(668. .1813)
                  /gene="dape"
                  complement(668. .1813)
                  /gene="dape"
                  /EC_number="3.5.1.18"
                  /note="NMA1730, dape, succinyl-diaminopimelate
                  desuccinylase, len: 381aa; similar to many eg. SW:P24176
                  (DAPE_ECOLI) dape, succinyl-diaminopimelate desuccinylase
                  from Escherichia coli (375 aa) fasta scores: E(): 0, 56.98
                  identity in 371 aa overlap. Contains Pfam match to entry
                  PF01546 Peptidase_M20."
                  /codon_start=1
                  /transl_table=11
                  /product="succinyl-diaminopimelate desuccinylase"
                  /protein_id="CAB84958.1"
                  /db_xref="GI:7380372"
                  /db_xref="SPTRMBL:09JUL0"
                  /translation="MTETQSLELAKALISRPVPPDRDCKLIVERLYKIGFAAEEL
                  HFQTKNIMWRGTVKPVVCFAGHTDVPTGPKVMDSPPEEPERDGRYGGAAAM
                  KTASCFVTAQERVAHPHQSIALITSDGDLDTGVVLYKAPGELIDIC
                  IVEPTAVDKLGMKNGRSGLSGNTLVKGGSHLAPHLAINPVRKPALELTO
                  EVMDGKNKYPPPTSFQISNTNGGANGNVPGLNKNFNRSTESTEALCKRVHAI
                  LDKRGVYDQWSSGQPTLTQAGKLTIDVAKAIAETGIEALSTTGGGSDRFKA
                  IAKELTLPSSNATIHQINENRDIPIKLSAYEGLIARLAGNAV"
misc_feature      complement(836. .1786)
                  /gene="dape"
                  /note="Pfam match to entry PF01546 Peptidase_M20, score
                  252.50, E-value 5.7e-72"
                  complement(1863. .1872)
misc_feature      /note="Core DNA uptake sequence: gccgtctgaa"
                  /label=DUS
                  complement(1927. .2547)
                  /gene="NMA1731"
                  complement(1927. .2547)
                  /gene="NMA1731"
                  /note="NMA1731, conserved hypothetical protein, len:
                  206aa; similar to SW:P44013 (Y552_HAEIN) hypothetical
                  protein from Haemophilus influenzae (207 aa) fasta scores:
                  E(): 0, 59.68 identity in 203 aa overlap."
                  /codon_start=1
                  /transl_table=11
                  /product="conserved hypothetical protein"
                  /protein_id="CAB84959.1"
                  /db_xref="GI:7380373"
                  /db_xref="SPTRMBL:09JUL0"
                  /translation="MLTPKSCDLFNIPFQFQSLTKKYQESIPQIKADYKNNQMOO
                  LIOVAALGAPAPAPPHIERKMGQVRAHFAFYRYEYQKNAALITLILNRRRISV
                  SLDMHCYKADVPRIALPDYNRMLDNDTEKYSFTDMHGHESEYIDYRTIVAQDNESDR
                  KLQNDDEFFCTIGHIERDGLGRODVAKWIAETVEDLLPLYEACHGK"
RBS               complement(2551. .2558)
                  complement(2570. .2579)
misc_feature      /note="Core DNA uptake sequence: gccgtctgaa"
                  /label=DUS
                  complement(2581. .3090)
                  /gene="NMA1732"
                  complement(2581. .3090)
                  /gene="NMA1732"
                  /note="NMA1732, conserved hypothetical protein, len:

```

```

169aa; similar to SW:P43971 (Y243_HAEIN) hypothetical
protein from Haemophilus influenzae (172 aa) fasta scores:
E(): 0, 49.7% identity in 169 aa overlap."
                  /codon_start=1
                  /transl_table=11
                  /product="conserved hypothetical protein"
                  /protein_id="CAB84960.1"
                  /db_xref="GI:7380374"
                  /db_xref="SPTRMBL:09JUL0"
                  /translation="MNPFTQSVTAEPLEMLYAGKVRFCGGVAMLSDYAENG
                  NOLVLOTIRQITRYFNVAAPLHHEDEENEPDLLQYAOQDSVDLLRQVYSLHGN
                  MDVAAEFAKLEADNAVIPDAEFKRFVAGYDVHIAIEEPLDMKMFIPKKEITIG
                  EIMARRRR"
                  complement(3185. .3736)
                  /gene="NMA1733"
                  complement(3185. .3736)
                  /gene="NMA1733"
                  /note="NMA1733, H.8 outer membrane protein, len: 183aa;
                  strongly similar to SW:P07212 (H8_NHME) H.8 outer
                  membrane protein from Neisseria meningitidis (183 aa)
                  fasta scores: E(): 0, 96.7% identity in 183 aa overlap.
                  Contains Pfam match to entry PF00127 copper-bind, Copper
                  binding proteins, plastocyanin/azurin family and Prosite
                  match to PS00196 Type-1 copper (blue) proteins signature."
                  /codon_start=1
                  /transl_table=11
                  /product="H.8 outer membrane protein"
                  /protein_id="CAB84961.1"
                  /db_xref="GI:7380375"
                  /db_xref="SWISS-PROT:P57025"
                  /translation="MKAYLALISAAYIGLAQSOEPAPAAEATPAAPASAPAAE
                  AAPADAAPAAAGNCAATVPSNDNQFNKDIQVACKKEFTITLKHGTQKTSKMGH
                  NIVIGTTEGMDGIFKRGDVGAAAPDYKPPDAVAVATKILGGSESSLTLPKALADG
                  EKFAKPTFGHGKILAMGKTYLD"
                  complement(3191. .3568)
                  /gene="NMA1733"
                  /note="Pfam match to entry PF00127 copper-bind, Copper
                  binding proteins, plastocyanin/azurin family, score
                  218.60, E-value 9.1e-62"
                  complement(3212. .3253)
                  /gene="NMA1733"
                  /note="PS00196 Type-1 copper (blue) proteins signature."
                  complement(4017. .4439)
                  /gene="NMA1734"
                  complement(4017. .4439)
                  /gene="NMA1734"
                  /note="NMA1734, unknown, len: 140aa"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical protein"
                  /protein_id="CAB84962.1"
                  /db_xref="GI:7380376"
                  /db_xref="SPTRMBL:09JUL0"
                  /translation="NMLQDLGRDALLSGLLKQAGQWRRLDAVAKLLPANLHPHQR
                  ACIEDGRVLTLAANNMAASRLKMIAPSVLPQAGIDASIRSVSLVLPKPKPTNT
                  IHLSKAALSEFSQSAVVKLEKRHDELALANLVRHGA"
                  4306. .4315
                  /note="Core DNA uptake sequence: gccgtctgaa"
                  /label=DUS
                  complement(4486. .4495)
                  /note="Core DNA uptake sequence: gccgtctgaa"
                  /label=DUS
                  complement(4570. .7320)
                  /gene="seca"
                  4570. .7320
                  /gene="seca"
                  /note="seca"
                  /note="NMA1735, seca, preprotein translocase SecA subunit,
                  len: 916aa; similar to many eg. SW:P10408 (SECA_ECOLI)
                  preprotein translocase SecA subunit from Escherichia coli
                  (901 aa) fasta scores: E(): 0, 58.0% identity in 920 aa
                  overlap. Contains Pfam match to entry PF01043
                  SecA_protein, SecA protein, amino terminal region and
                  Prosite match to PS01312 Protein seca signatures."

```





```

/codon_start=1
/transl_table=11
/product="acetylactate synthase, large subunit (llyb-3)"
/protein_id="AAB89235.1"
/db_xref="GI:2648516"
/transl_table=11
/product="MDGA5AAKALLREKIDCA5LPSGELLPVCEYLAEAGVRIET
RHEQAGNADGAYAVTRRPGFCIVPTGGLANLMPALQAQVYASSPVLAJAGISRYV
NIDRAFEEDIDGVLWKEVYTKWMSRLVPTTHRLHEVYDCAFRALSGRYGVLLEPKD
VLMADIDKVELAEPRNRYRTGVCCEEEYIRALEMLINAEKPLIYAGSGVYASRAE
NEMIADEKELISPVSGIGGCCTISDPHPLFAGQASAPAMQSDPVIYVTRDEEL
GEGCKKGVVHVDIDACELAKNRVDAISCDAGYFSSKALAMEGKKEENARPI
MEGVNRFEEAKGEDKPKPQRLMRREINELVSRDVIIDGGETTAGLLYLKRGV
ISSQGFEGHLAGIPMGIAAKAAYPEKRFVYVTDGDFLFGAGIEDTAARHGLVVVY
VNDSAMGLVNHTRILSTKSAERASYGYMLNPARAYDRFAESLAEYGEVSEKVE
AVKRAFDPALPAVDVRYKIDELISPLAYILSGSFG"
2909..3832
/gene="AF2016"
2909..3832
/gene="AF2016"
/notes="similar to GB:L77117 SP:Q57664 PID:1590951 percent
identity: 30.00; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="UDP-glucose 4-epimerase (galE-2)"
/protein_id="AAB89234.1"
/db_xref="GI:2648515"
/transl_table=11
/product="MAVLITGAGFLGRFVAKHPSVILDLKKGKGFEGGSVTAM
SDIAEVPKHEIEGVIAHAELSVKAEKSHDARAVEGALNLEACRQDQKVF
TSSHYGPRSPYPTFEFSYRDPTEFGATKACGSLITTYSTYTGIDFRAVRPILVG
PFRMGASVAFSLIDAPENGSTVITLPEKRLPVLYKDAELVRLSIDRVSK
PIRVGVVLSIKDIVERAKKFTIPKQFRIIDDSRRIADQWILMTVERKAGIIDKY
SRLELDEIMKSAEELVEDHKTAEVEE"
3829..4674
/gene="AF2017"
3829..4674
/gene="AF2017"
/notes="similar to GB:M31799 PID:144733 percent identity:
35.36; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="3-hydroxyacyl-CoA dehydrogenase (hbd-9)"
/protein_id="AAB89233.1"
/db_xref="GI:2648514"
/transl_table=11
/product="MKYFVIGAGLNGGIAIAASKHEVVLQVSEKALFANEQIPE
ELSKIEPTTLEKVDKCDIVMEAVFEDLMTKVEVLEKVEBRLTNAKCSNYSIVDD
IAERLDSPSRLGVMNMPHVPMLVEIVISRFDSKTVAFVEGFLTELGEVYVKG
QSLVNRNNAVLSEASRMIEGVAEVDVWKHGLITLFLGPGINDLYIGIDVAY
YASLYIKRRGDEKRPPEMLQEKIKKGEVGAQKGYEYGPRAVERERLKLRL
FLGLE"
4671..5405
/gene="AF2018"
4671..5405
/gene="AF2018"
/notes="similar to GP:454844 percent identity: 26.98;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB89232.1"
/db_xref="GI:2648513"
/transl_table=11
/product="MNEATVAMFLITLIVAFVTVVVGSGSAILSPISNMPREKA
VVLIILESFVNYFLLKERNFSLEIYIGAFPGIAGLIFGLLPQELVGLTIGVA
MGYMAVMELGVFRVARTKRLFLSKGFLSGAGVITGVGDPVITLLNQGVAASFI
RSFMTYLVIVDTATLISAFPAFGHVNQVLIINPLILAPSIILAYALGSRILSMRDE
LKRIILSAVLLSVILRYGGLIG"
5422..6537
/gene="AF2019"
5422..6537
/gene="AF2019"
/notes="similar to SP:P45513 PID:493086 percent identity:
35.71; identified by sequence similarity; putative"
/codon_start=1

```

```

/transl_table=11
/product="alcohol dehydrogenase, iron-containing"
/protein_id="AAB89231.1"
/db_xref="GI:2648512"
/transl_table=11
/product="MRVFGMGRLEERIEEVERLKGVYFIADRAVRRAVEELNSLF
DYEFSDVEPEPEIGIVDDIIEKFSPTDVVGSGSTLDVAKTLTAACAEKAVELY
GKKIPEKCKILLIPTAGGSSVTAAVRIKREKRYVFEDESILADIATIDPKLT
LSCPHYATSGGIDAFACATLEATYSLSNFTSDMEFAKAIATIIPKALREYANGENV
ARTDMFASLILAGIENSAGTGALGHALGFAHNHIDMSPHKSVGIMPVYLOYNATAD
LEKHAIRARNMPVVRREEMDVVRRKSIIHGLIREEY"
complement(6502..7077)
/gene="AF2020"
complement(6502..7077)
/notes="hypothetical protein; identified by Genemark;
putative"
/codon_start=1
/transl_table=11
/product="A. fulgidus predicted coding region AF2020"
/protein_id="AAB89241.1"
/db_xref="GI:2648523"
/transl_table=11
/product="MWKIRLHTVTSSETRAKIEDLKRRHTTSVVERKAVDELTYSEN
FSRLGDEDLILAFIREINELMCADKHDTALVGDRAVRRESMIAVAYLSKPKIS
DLDFEELSVARLNMNLNBAEHAEOXKGEKINFVYHDMRSKAVSELHMLKLYL
EKYSSKYEOMQVDTIIVNGFSVLEFPKSDVD"
7057..8148
/gene="AF2021"
7057..8148
/gene="AF2021"
/notes="similar to GB:D37799 SP:P39763 PID:520840
PID:1377825 GB:AL009126 percent identity: 26.60;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="rod shape-determining protein (mreB)"
/protein_id="AAB89230.1"
/db_xref="GI:2648511"
/transl_table=11
/product="MEPDYHGFPYPTPLKTYPEEIFYGKCTENMOPVGLDIGNYT
KITADKNVLIVSYAVAGEKMSLKGETKSVYICDEALISVQSMENVEVRLPHBG
RVLHTSVFLAKYGLKLLKVKDGLITGLPYSSKREBELRELESLNSVLLFEPE
PVGTMAVMGMEVGCVDIGFTDVLVADMEYLGKDTMLMGVLEKYLEVAVNRI
GLSITPEEMTKLITTEGYEIGIRGKRVSVRELEIMPEYEVTVTWERIASRVKLV
LEGLSTPEENALITGGGSMPLGVYEMFEESDIDGVKRPDEPIKANSIGYIIAKS
LVEAETEGSGAEIILDADKKKGRKK"
8145..8672
/gene="AF2022"
8145..8672
/gene="AF2022"
/notes="hypothetical protein; identified by Genemark;
putative"

```

```

Query Match      8 1%; Score 86; DB 1; Length 22014;
Best Local Similarity 47.4%; Pred. No. 5,5e-13;
Matches 476; Conservative 0; Mismatches 495; Indels 33; Gaps 6;

```



```

metabolism; 2-prime-deoxyribonucleotide metabolism"
/notes="Product confidence : putative
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PUTATIVE TRANSMEMBRANE RIBONUCLEOSIDE REDUCTASE
(SMALL CHAIN) OXIDOREDUCTASE PROTEIN"
/protein_id="CAD16511.1"
/db_xref="GI:17429827"
/translation="MLSWDDVKKPAAQQAQPAQVOPQIPQSQPTDQGGALPSSAT
QAAGILGNPNNAQAQSSRVNAADKRINGATDVNOLVFPKYKNAWEKYLACANHW
MPOEINNSRDIALKDPNGLTEDERLIKRNGLFFVTADSLAANNIVIGYKQITAPE
CROYLLROAFEEAIHHAQYIVESLGLNEALFNAYHEVQSIRKQDEFLIPDITLT
DPSFKGTENQOKLLSLIVACIMEGLFFVYGETQILAMGRQNKMTGAABQYQYIL
RDSLCNFGIDILQIKLENPHLTAFAEKITELEFKAVDLEYAEDTMRGVLG
LNAPMFKYLRFICNRRCQOIGLDALFPNEDNPFPMWSEMDLKKERNFFETRVIEQ
TGALSWD"
complement(3477..6440)
/gene="nrda"
/notes="nrda"
/notes="RSC2805; RS00300"
complement(3477..6440)
/gene="nrda"
/EC_number="1.17.4.1"
/function="small molecule metabolism; central intermediary
metabolism; 2-prime-deoxyribonucleotide metabolism"
/notes="Product confidence : putative
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PUTATIVE RIBONUCLEOSIDE REDUCTASE 1 (LARGE
CHAIN) OXIDOREDUCTASE PROTEIN"
/protein_id="CAD16512.1"
/db_xref="GI:17429828"
/translation="NOTAQTEHSGTANTPAFSQPAECTAVASGAGYADYKLIIRNGA
VVFEPKSIAMVMTKFLAVNGGQGAASARVELVEGLTONVVRALLSRPNNGTFHI
EDVQVQELALMRSEHVDARAYVLYRERROERAQAGETQOQAFILGNVTDGGITR
PDMAALNRLIVSACAGSVDPEPIKETVKNLYEGVPMQSVQVDSIAAARTLIEK
DPAYSOATARILLHTIRREILGEVTOAEMAGRYADYFPQIKRGINAGLLDDTLAOF
DLAKGALDASRDLOFNVLQTLVDRYFLHISDKRIEMPOAFEMRMVAGLSLEID
REARAEFYQLLSRSDNMSPTLFNSCTCRSOLSSCYLTVSDDLGLIYEALKENAL
LSKFAAGLGNQWTVNRALGSHIKGTNGKSQGVVFLKVVNDTAVAVNOGGKAGVCA
YLETHWLDIEFLELRKNGTDDRRRTHDMTANWIPDLFMKRVMEGGEWTLFSPSTCP
DLHDKVGRFAEAYLYGSEDKVARGEIKLFKKMPALQMLRMLFETGHPWITFKDP
CNIRSPQOHVGVHSSNLCETITLNTNDSAVNGLSVNLVAHVVKQADGSYALDHD
KIKKTVRTAMRLDNV IDINTYAVKKARDSNLRHPVGLGIMGFQDALHVLRIPYAND
AAVQFADTSMQAVCYAYWASTELAEERGVSTYKGSISMDRGVLPDPSLKLAEERG
YLEADLSMTMDGLRERIOOHGRMNSVIAIATATISNIGVSACEPTFYQNLVYK
SNLSEFTVNDVLRDLKARGLDNEVADLKYFDGSLARIDRIPDRLRLYATAFE
VHPQWLVEASRQRKRWIDQASLNIMAGASGKMLDPTYKLAWLRLKTYLYLTIGA
THPERKTSRGTLNVAVSSGSDAGSVAAGRAPAPASALDTPAATAPMPDAEGAVCT
MRPDGPFEECEACQ"
complement(6877..7479)
/gene="ampb"
/notes="RSC2806; RS00299"
complement(6877..7479)
/gene="ampb"
/function="cell processes; protection responses;
detoxification"
/notes="Product confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11

```

```

/product="PUTATIVE AMIDASE PROTEIN"
/protein_id="CAD16513.1"
/db_xref="GI:17429829"
/translation="MPEVAGPVIRIGADGWADGVCRHPSNVNDARPEGLPIDLIVLHNI
SLPPARFGTDDVLDFFANTLDCGGAHYFEQURGVRSAAHFFIRRTGACVQAPCDARA
WHAGVSDFRGTNRDNDSIGIEIGTDDPPTPAYATAALVRAICAAYPIRALAGH
SDIAPGRKTDPGPCDFDWAHLRLAKLEAALFPYQAGAAPA"
complement(7472..9184)
/gene="pehr"
/notes="RSC2807; RS00298"
complement(7472..9184)
/gene="pehr"
/function="small molecule metabolism; global functions;
global regulatory functions"
/notes="Product confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE RESPONSE REGULATOR TRANSCRIPTION
REGULATOR PROTEIN"
/protein_id="CAD16514.1"
/db_xref="GI:17429830"
/translation="MSRQPPFRPPMSKAAVVREPILVDDDEADRELELEISLRMGHD
VVLGGLAEAREALTRORFALVLDMLRGDLGLDLVRLQSLADRTPVAVITAYGSA
ENAVELKAGAFDYIAKPLSDQLSLVLNALGRQDDPDGSDADLAERTSALLPGHS
AAMOEVRSLRLRLARSAPVVISGESGKERAARAIALHSRSRARPVAVNCGAIEPE
NLMEAEFFGVYKGAFTGADSDROGFFQAAHGGTLMDELVDADPLTMQVKLLRAQERR
VRKIGESREDPDVVRVVCASHONLARLVAAAGRFREDLFYRLNLELMPILREAREDV
FVLAGVLLEQLAARYGDLRPLKRLTRQALQQLCATYFFGNVRELDNLLERAYFAEGES
IDVDHLGAUGTIDIBRSFLHRAEHGGHPVPAHLPVPAPGHPAHPGHPGHVHAHPL
GVPQVGGWPDAAIYIPVPMGLMPHVVPVPEPAPVPEAPAMPAPVSLPVDLPA
YLESVERGVILAAQAQTFNRATASKLLGLSFLRQYRMQQLDIRDPRDIEAAGNGG

```

Query Match

7.6%; Score 80.6; DB 1; Length 188050;

Best Local Similarity 46.2%; Pred. No. 2.3e-11;

Matches 475; Conservative 0; Mismatches 524; Indels 30; Gaps 5;

```

QY 26 TTGAAGCTTACAAAGCAGGAGTCACTCCCGCTCCGTCAGCGCTTTCCTTCTTAACGAATTC 85
DB 168373 TGGCGCGCTACACCTGCCGACGCTCCCGCATGGTCAACCTGACGCGATGAGAAC 168314
QY 86 CCTACGACTTTCCCGAGGAGATAAAACAAGGGCTTAGAAGAATTAATAAAGGTTCCTC 145
DB 168313 CCTACCATCTCGCGGCCATGCTGCGCGACGCTCGGCAAGCGCTGCGCGAGTGGCG 168254
QY 146 TCAACAAATACCCAGACCCCGCAAGAGATTAAAGCGGTTCTTTCGGGATTTTTCG 205
DB 168253 TGAACCGCTACCCGTCGCCACCCCGCCGCTCAGGCCCGACCTCAAGCGCGTGATGC 168194
QY 206 CGGTTAAG---GAAGAAATTTAGTTCTTCGTAACGGTTCGGACCACTCATATATAC 262
DB 168193 CGGTGCGCGCGCGCGCGACGTGCTGTGGCAACGGCTCGGACGAGATCATACGCTGA 168134
QY 263 TCTCAATAGCTATAGTGAATTTTACATACCCGTTTACATACCTGTTCCCACTTTCCCA 322
DB 168133 TCGGTATCGCGCGCGCAAGCCCGCGCGCGCTGCTGGCGCGGTCGCGCGGCTTCGTGA 168074
QY 323 TGTACGAGATAAGTTCGAAAGTTCTTCGGAAGACCCCTCGTAAAGGTTCAACTGACGAAA 382
DB 168073 TGTATCGATCTCGCGCGCAGCTGATGGGTGGAAGTTCTGTTGGCGTCCCGCTCAGGCCG 168014
QY 383 ACTTTGATATAGACTTAGAAGAAGTATTGAATTAATAGAGAAGAAAAACCGTTCTCG 442
DB 168013 ATCTCAACCTCGACCGCGAAGCGATGCTGCGCGCATGGTGTAGACACCGCGCGCTCA 167954
QY 443 GGTACTTTGCTTACCCCAACACCCACGGAACCTCTTTTCCAGGGGAAAGATTGAGG 502
DB 167953 TCTACCTGGCTATCCGAACAACCCGCGCAACCTGTTTCGACGCGCGCGCATGGATG 167894
QY 503 AGATAAGAAACAGGGGTGTTT-----TCTGTGTAATAGACGAAGCCT 544

```



[illegible]

```

ILGYAIEKNGDVVPLEGRNISDFNIGEKIRSHTPPHVIEAEAEKLEAMSQAERFEFFWK
EFAKICIKYACRQVCPMCYCRRCIVDNVQNPQWISTSHTLGNEFWNLVRAFLHAGRC
AACGACDRACPNVPLRLNLYRMGKEVRSFADYVAGENSODKPPVLASFQDDPETFIL
"
6005..7018
/gene="hydB-2"
/note="CT1249"
6005..7018
/gene="hydB-2"
/codon_start=1
/transl_table=11
/product="hydrogenase/sulfur reductase, beta subunit"
/protein_id="AAM72479.1"
/db_xref="GI:21647225"
/translation="MRLILRLKDECLAAQKAGFVLAIPVVKRHEMSCFGEVOKSGD
MALDILVETRIKIDFFOTEPTELIKIRIGKQIDSETMTPEKKRVFVGRPCDASGL
AIDDLFGWVDYKDDYFRRREKSVIIVTIAQADDFCMCTSVKLSPOSTKGADWMLRP
LSDGSGWDAVSRDGTALVDVTSLLQESAAPVQVAEKDFDVEKMWGLADKEN
FSQFWKIDALRCVCGSCTFLCPTCHGFDIQDEGDTYQGIIRKKNWDSCEPFLTMHT
SGHNPRNTQSTWRORIMHKFNRYRKGFGVNSCSGCGRCRQCPCVDMGITETLQAITN
LPR"
7045..7881
/gene="hydB-2"
/note="CT1250"
7045..7881
/gene="hydB-2"
/note="identified by match to PFAM protein family HMM
PF00175"
/codon_start=1
/transl_table=11
/product="hydrogenase/sulfur reductase, gamma subunit"
/protein_id="AAM72480.1"
/db_xref="GI:21647226"
/translation="MYSFPMRVSVKRAEAPGVNTLKLFEVKQEDHEFEKANYRTGM
FGLYVFGGESFTCVASPETREKYEICTFRSGRVTSTLANTDAGDIVTPRGPKGNR
FPIEEPENLLFIAGGIALPPTSRVNSCLDQREKIRYDVTIVIGARIVADLVYKNEL
DEMKQDRLVLTVPDGGETPDQDHFVFTVLEQAAPSPENTIAVLGCGPIMIKF
LTIALEKLGTAENVYTTLENRMKGIGKRCNRSVYICKEGPVFTAAAEVQAMPQAA
DL"
7887..8000
/gene="CT1251"
7887..8000
/note="CT1251"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM72481.1"
/db_xref="GI:21647227"
/translation="MKMDTKRGRKRSFLFLCKSLFLIHLHDPHPNIIILIFRHH"
8028..8591
/gene="pth"
/note="CT1252"
8028..8591
/gene="pth"
/note="identified by match to PFAM protein family HMM
PF01195"
/codon_start=1
/transl_table=11
/product="peptidyl-L-trNA hydrolase"
/protein_id="AAM72482.1"
/db_xref="GI:21647228"

```

```

Query Match      6.5%; Score 68.8; DB 1; Length 11265;
Best Local Similarity 45.4%; Pred. No. 3.8e-08;
Matches 447; Conservative 0; Mismatches 522; Indels 15; Gaps 5;

QY 61 GTCAGGCTTTCTCTACGAATTCCTCCAGAGATTTCCCGAGGAGATATAACAAAGGGCC 120
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 11070 GTCAGGCTGAACGAGAGAGAACCCCTTCGATCTGCCCTCGTGGCTCAAGGACGAAGATT 11011
QY 121 TTAGAAGATTAAGAGTTCCCTTGACAAATACCCAGACCCCGAAGAGAGCTTA 180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 11010 CTCGACCAGTTCCGCCATGAGCCCTCGAACCCTATCCGGATATTTCCGGTACCAGGGGC 10951
QY 181 AAACGGGTTCTTGGGGATTTTTCGGCGTAAAGGAAGAAAAATTTAGTCTCGGTGAACGGT 240
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10950 ATGGCGGCTATGTCATCGTTTCTCGGCGTGAAGCCGGAGTTGGTGATCATGACGAATGGT 10891
QY 241 TCGGACGAACCTCATATCTACTCTCAATAGCTATAGGTGAACTTTACATACCGCTTTAC 300
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10890 TCGAACGAGATGCTCTACACCATCTTCATGSCCTCGCTCGGCGCTGGTGCAGAGTGCTT 10831
QY 301 ATACCTGTTCACCACTTTCCTCCATGATGAGATAAGTGGAAAGTTTCTCGGAAGACCCCTC 360
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10830 ATTCCAGAGCCTTCGTTTTCGCTTTACGACAAAGCTTCGACGCTTTCACAGAGGCTGGGGTT 10771
QY 361 GTAAGGTTCAACTGGAGCAAAACCTTTGATATAGACTTAGAAAGAAAGTATTGAATTAATA 420
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10770 GTCAGAGTGCCGATGATGACGATCTCTCCTTTGACGTGCGATCGATCATCGAAGCCGCC 10711
QY 421 GAGAAAGAAAAACCCGTTTCTCGGCTACTTTGCTTACCCAAACACCCACCGGGAACCTC 480
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10710 AGGCGGAGAGAGGTGCGATTTTATCGTCTCCACGCGCAACAATCCGACGACGAAGTCG 10651
QY 481 TTTTCCAGGGGAAAGATTGAGGAGATAGAAACAGGGGTGTTTCTGTG---TAATAGAC 537
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10650 CTGTCCACGAGATGCGAAGCATTTGCGAAGCTGCCGAGCCATTTGCTGCTGTCAT 10591
QY 538 GAAGCTACTATCAATTACTCCGGAGAAACCTTTCTTGGAAAGACGCGCTCAAAAGGGAAGAT 597
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10590 GAAGCGTATGTCGAGTTTTCGCGTGAGCAATCGCGCTCGACCTGATCGACGCTACCCG 10531
QY 598 A---CGGTAGTTTTCGAGACACTTTCAA---AATCGGTATGGCGAGTTTAAAGGTAGGG 651
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10530 AACCTGATCGTGTTCGCGACCATCTCCAAAGCGCTTTCGCGTGGCGCGGAATCGGTATCGGC 10471
QY 652 ATTTTAAATAGGAGGGGGAATCGTCTCAGAAATTAACAAGGTGAGACTCCCTTCAAC 711
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10470 TTTCGATTTGCAATCCGAGCTGCTGCGCCGAATCTCCAAGCCGAAGATTCCCTTTCGCG 10411
QY 712 GTGACCTACCCCTCTCAGGTGATGGCAAAAGTTCTCTCAGCGAGGGAAGAGATTCCTTA 771
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10410 TCGAGCGGACTGGCGGAATCACTCTCATGCTGTGCTC---GAAAACTACTCACTCGTG 10354
QY 772 ATGGAAAGATACAGGAGGTGTTACAGAGCGAGAAAGATGTACGACGAAATGAAGAAA 831
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10353 ACCGACGCGTGCATGATACCTTCGCCGAGCGCGGCCGATCGAGGCGGAGCTGACAAA 10294
QY 832 ATAGAAGAGTTTGAGTGTTCCTCCGAGTAAGGCTAACTTCTTCTTTTTCAGAACGCTTAC 891
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10293 ATTCTGGCATCCATACCTTCGAGAGCGACACCAACTTCTCTCATCTTCGCTGCCAAC 10234
QY 892 CCCGCCACGAGGTTTATCAGGAGTACTGAAAAGGGATGCTCTCGTCAGGACGATATCT 951
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10233 GCATC---CGAGGTGTTTCAGGAAGCTCAAAATGCGGAGTCTGGTTTCGGAACGTGCG 10177
QY 952 TACATGGAGGACTCCAAAAGTCCCTCAGGTAAAGCTAGGGAACCCGGAAGAAACAAAC 1011
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10176 GGATACCCGCTTATGGAGAACTGCTCCCTCCGCTCAATGTTGGTCTCAGGAGGAGACGAC 10117
QY 1012 AAGTTTCTGGAAGCACTGGAGGAG 1035
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||
DB 10116 CGCCTGTTGGAACCTGCTGAAAAAG 10093
      ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||  ||||| |||

RESULT 14
AY117255
LOCUS
DEFINITION
Arabidopsis thaliana putative histidinol-phosphate aminotransferase
(AT5g10330) mRNA, complete cds.
ACCESSION
AY117255.1 GI:21436320
VERSION
FLJ_CDNA.
KEYWORDS
thale cress.
SOURCE
Arabidopsis thaliana
ORGANISM
Arabidopsis thaliana

```







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 13:04:46 : Search time 302 Seconds  
(without alignments)  
7941.650 Million cell updates/sec

Title: US-09-905-173-23

Perfect score: 1065

Sequence: 1 atgataccaccagaggattaa.....ccctttcaagctctcttttaa 1065

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

```

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1065	100.0	1065	18 AAT78778	Aquifex histidinol
2	90.8	8.5	13732	21 AA81484	N. meningitidis pa
3	90.8	8.5	34980	21 AAF21611	Neisseria meningit
4	90.8	8.5	837096	21 AA81489	N. meningitidis pa
5	60.2	5.7	1664976	19 AAV21209	Methanococcus jann
6	46.2	4.3	1085	24 ABK73007	Bacillus lichenifo
7	42.2	4.0	1092	18 AAT78780	Ammonifex histidin
8	37.6	3.5	588	22 ABA59603	Human foetal liver
9	37.6	3.5	588	22 ABA28180	Probe #6646 for ge

10	37.6	3.5	588	22	AAK07876	Human brain expres
11	37.6	3.5	588	22	AAK33743	Human bone marrow
12	37.6	3.5	588	22	AAI16517	Probe #6450 for ge
13	37.6	3.5	588	22	AAI39468	Probe #8154 used t
14	37.6	3.5	588	24	ABS08616	Human genome-deriv
15	36.8	3.5	9835	23	ABL21524	Drosophila melanog
16	36.4	3.4	6478	22	AA543416	Chemically pretrea
17	36.4	3.4	6478	24	ABN80200	Human chemically m
18	36.4	3.4	6626	22	AA546809	DNA transcription
19	36.4	3.4	6626	22	AA546809	Tumour suppressor
20	36.2	3.4	6804	18	AAV74676	Staphylococcus aur
21	35.8	3.4	560	24	ABK77425	Bacillus clausii g
22	35.8	3.4	1098	22	AAH67263	C glutamicum codin
23	35.8	3.4	1221	22	AAF71952	Corynebacterium gl
24	35.8	3.4	10732	21	AAAF1998	Corynebacterium gl
25	35.8	3.4	10732	21	AAAF1998	Gene encoding a su
26	35.8	3.4	34980	22	AAH68531	C glutamicum codin
27	35.4	3.3	4590	22	AAH24065	Yeast AOD9604-asso
28	35.2	3.3	1104	24	ABO86187	Lactobacillus rham
29	35.2	3.3	1350	24	ABD31862	Lactobacillus rham
30	35	3.3	5621	24	ABL32205	Human immune syste
31	35	3.3	24187	23	AA59529	Propionibacterium
32	34.8	3.3	647	24	ABQ68411	Listeria monocytog
33	34.8	3.3	4213	24	ABQ70936	Listeria monocytog
34	34.8	3.3	6631	22	ABA19734	Human nervous syst
35	34.6	3.3	6631	22	AAK84213	Human immune/haema
36	34.6	3.2	263	23	ABV61754	Human prostate exp
37	34.6	3.2	1341	22	AAI93087	Human polynucleoti
38	34.6	3.2	2483	24	ABK84870	Nematode infectio
39	34.6	3.2	8539	24	ABK84868	Nematode resistanc
40	34.6	3.2	19243	22	AAK71662	Human immune/haema
41	34.4	3.2	375	22	AAK06304	Human brain expres
42	34.4	3.2	375	22	AAK31971	Human bone marrow
43	34.4	3.2	375	24	ABS06734	Human genome-deriv
44	34.4	3.2	12225	18	AAT89257	Human bg gene (sho
45	34.4	3.2	12616	18	AAT89255	Human bg gene (lon

#### ALIGNMENTS

RESULT 1  
AAT78778  
ID AAT78778 standard; DNA; 1065 BP.  
AC AAT78778;

DT 18-FEB-1998 (first entry)

DE Aquifex histidinol-phosphate aminotransferase VF5/HPA DNA.

XX Histidinol-phosphate aminotransferase; VF5/HPA; chiral compound;  
KW ss.

XX Aquifex sp. strain VF5.

PN WO9729187-A1.

PD 14-AUG-1997.

PF 21-JAN-1997; 97WO-US01094.

PR 08-MAY-1996; 96US-0646590.

PR 09-FEB-1996; 96US-0599171.

XX (RECO-) RECOMBINANT BIOCATALYSIS INC.

XX Swanson RV, Warren PV;

XX WPI; 1997-415343/38.

XX P-PSDB; AAW24254.

XX New transaminase(s) and aminotransferase(s) derived from host cells

PT - used for producing enzymes, and hybridisation probes for a cDNA or genomic library

XX Claim 1; Fig 7; 95pp; English.

XX This genomic DNA encodes histidinol-phosphate aminotransferase V<sub>P</sub>5/H<sub>P</sub>A (see AW24254) of Aquifex VF5, a strictly chemolithoautotrophic, marine eubacterium which grows optimally at 85-90 deg C and pH 6.8 in high salt medium. The V<sub>P</sub>5/H<sub>P</sub>A sequence can be amplified from a pBluescript vector that contains the DNA by PCR (see AAF78794-95). The DNA can be used for recombinant production of the aminotransferase and to develop hybridisation probes. Claimed thermostable transaminases and aminotransferases (AAW24248-57) can be produced from native or recombinant host cells for use with L- and/or D-amino acids for production of optically pure chiral compounds used in the pharmaceutical, agricultural and other industries. A method is claimed for transferring an amino acid group from an amino acid to an alpha-keto acid using a claimed enzyme.

XX Sequence 1065 BP; 344 A; 220 C; 252 G; 249 T; 0 other;

Query Match 100.0%; Score 1065; DB 18; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATACCCAGAGGATTAAGGAACCTTGAAGCTTACAAGACGAGGTCACTCCCGCTCC 60  
DB 1 ATGATACCCAGAGGATTAAGGAACCTTGAAGCTTACAAGACGAGGTCACTCCCGCTCC 60  
QY 61 GTCAGGCTTCTCTAAGCAATTCCTTACGACTTTCCCGAGGAGATATAAACAAGGCC 120  
DB 61 GTCAGGCTTCTCTAAGCAATTCCTTACGACTTTCCCGAGGAGATATAAACAAGGCC 120  
QY 121 TTAGAAGAAATTAAGGCTTCCCTTCAACAATATACCCAGACCCCGAAGGAAAGAGTTA 180  
DB 121 TTAGAAGAAATTAAGGCTTCCCTTCAACAATATACCCAGACCCCGAAGGAAAGAGTTA 180  
QY 181 AAAGCGGTTCTCGGATTTTTCGGGCTTAAGGAAGAAATTTAGTCTCGGTAACGGT 240  
DB 181 AAAGCGGTTCTCGGATTTTTCGGGCTTAAGGAAGAAATTTAGTCTCGGTAACGGT 240  
QY 241 TCGGACGAACATATACCTCTCAATAGCTATAGTGAACCTTTACATACCCGTTTAC 300  
DB 241 TCGGACGAACATATACCTCTCAATAGCTATAGTGAACCTTTACATACCCGTTTAC 300  
QY 301 ATACCTGTCCACCTTTCCCATCTACGAGATAAGTCGGAAGTTCTCGGAAGACCCCTC 360  
DB 301 ATACCTGTCCACCTTTCCCATCTACGAGATAAGTCGGAAGTTCTCGGAAGACCCCTC 360  
QY 361 GTAAAGGTTCAACTGGAGAAACCTTTGATATAGACTTAGAAGAAGTATTGAATTAATA 420  
DB 361 GTAAAGGTTCAACTGGAGAAACCTTTGATATAGACTTAGAAGAAGTATTGAATTAATA 420  
QY 421 GAGAAGAAACCCGTTCTCGGTAACCTTTGCTTACCCAAACACCCACGGAACCTC 480  
DB 421 GAGAAGAAACCCGTTCTCGGTAACCTTTGCTTACCCAAACACCCACGGAACCTC 480  
QY 481 TTTTCCAGGGGAAGATTGAGGAGATAAGAAACAGGGGTGTTTCTGTGTAATAGACGAA 540  
DB 481 TTTTCCAGGGGAAGATTGAGGAGATAAGAAACAGGGGTGTTTCTGTGTAATAGACGAA 540  
QY 541 GCCTACTATCATCTCCGGAGAACCTTTCTGGAAGACGGCTCAAAAGGAAGATACG 600  
DB 541 GCCTACTATCATCTCCGGAGAACCTTTCTGGAAGACGGCTCAAAAGGAAGATACG 600  
QY 601 GTAGTTTTCAGGACACTTTCAAAATCGGTATGCGAGCTTTTAAGGGTAGGGATTTTAATA 660  
DB 601 GTAGTTTTCAGGACACTTTCAAAATCGGTATGCGAGCTTTTAAGGGTAGGGATTTTAATA 660  
QY 661 GGAAGAGGGGAAATCGTCTCAAGAAATTAACAAGGTGAGACTCCCTTCAACGTGACCTAC 720  
DB 661 GGAAGAGGGGAAATCGTCTCAAGAAATTAACAAGGTGAGACTCCCTTCAACGTGACCTAC 720

QY 721 CCCTCTCAGGTGATGGCAAAAGTTCTCTCACGGAGGAGAGAAATTCCTAATCGAAAG 780  
DB 721 CCCTCTCAGGTGATGGCAAAAGTTCTCTCACGGAGGAGAGAAATTCCTAATCGAAAG 780  
QY 781 ATACAGGAGGTTGTAACAGACGAGAAAGATGTACACGAAATGAAGAAATAGAAGGA 840  
DB 781 ATACAGGAGGTTGTAACAGACGAGAAAGATGTACACGAAATGAAGAAATAGAAGGA 840  
QY 841 GTTAGGTTTTCAGTAAGCTTCTTCTTTCAGAACGCTTACCCGCCCCAC 900  
DB 841 GTTAGGTTTTCAGTAAGCTTCTTCTTTCAGAACGCTTACCCGCCCCAC 900  
QY 901 GAGGTTTATCAGGAGCTACTGAAAGGGATGCTCTCTCAGGAACGATCTTACATGAA 960  
DB 901 GAGGTTTATCAGGAGCTACTGAAAGGGATGCTCTCTCAGGAACGATCTTACATGAA 960  
QY 961 GGACTCCAAAAGTGCTCAGGTAAGCGTAGGAAACCGGAAGAAACAAAGTTCTG 1020  
DB 961 GGACTCCAAAAGTGCTCAGGTAAGCGTAGGAAACCGGAAGAAACAAAGTTCTG 1020  
QY 1021 GAAGCACTGGAGAGATATAAATCCCTTTCAAGCTCTCTTTAA 1065  
DB 1021 GAAGCACTGGAGAGATATAAATCCCTTTCAAGCTCTCTTTAA 1065

RESULT 2

AAA81484

ID AAA81484 standard; DNA; 13732 BP.

XX AC AAA81484;

XX DT 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm\_32 SEQ ID NO:32.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
KW Meningococcus B; MenB; ds.

XX OS Neisseria meningitidis.

XX PN WO200022430-A2.

XX PD 20-APR-2000.

XX PF 08-OCT-1999; 99WO-US23573.

XX PR 09-OCT-1998; 98US-0103794.

XX PR 30-APR-1999; 99US-0132068.

XX PA (CHIR ) CHIRON CORP.

XX PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC; Scarlato V;

XX PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Rappuoli R, Pizza M;

XX PI WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be used in the diagnosis and treatment of N. meningitidis infection and other Neisserial infections, for example, N.gonorrhoea -

XX Claim 7; Page 589-593; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic proteins from Neisseria genomic sequences. AAA81453 to AAA82414 represent specifically claimed Neisseria meningitidis genomic DNA sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent Neisseria DNA sequences and their corresponding proteins; AAA81254 to AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the isolation of Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF



QY 242 CGGACGAACTCATATACCTCTCAATAGCTATAGGTCAACTTTTACATACACCGTTTACA 301  
 DB 144679 CGGACGAACTCATATACCTCTCAATAGCTATAGGTCAACTTTTACATACACCGTTTACA 144738  
 QY 302 TACCTGTTCCCACTTCCCATGTACGAGATAAGTGGCAAGATTTCTCGGAAGACCCCTCG 361  
 DB 144739 CAGCGCAACCGATTTCGTCATGTACCGCCACACCGCGCTGTACGCGATGGATTATG 144798  
 QY 362 TAAAGGTTCACTGACGAGAACTTTGATATAGACTTAGAAGAAGTATTGAATTAATAG 421  
 DB 144799 TCGGCTTCCACTGACGAGAGATTTTCACTCAACTGCGCCGCTCTCGAAGCGGTCA 144858  
 QY 422 AGAAGAAACCCGTTCTCGGTACTTTTGTCTTACCAAAACACCCCGGGAACCTCT 481  
 DB 144859 GGAACACCGCCTGCTGCTGACTTTATCGCTACCCCAACACCCCGCGGTATGCT 144918  
 QY 482 TTTCCAGGGAAGATTTGA---GGAGATAAGAAACAGGGGTGTTTCTGTGTAAATAGACG 538  
 DB 144919 TCACGCGTCCGAAATCGAAGCGCTCATCGAAGCTTCAGACGGCATCGTCTGTCGTGATG 144978  
 QY 539 AGCCTACTATCATTTACTCCGGAGAACCTTTCTGGAAGACGCGCTCAAAAGGAAGATA 598  
 DB 144979 AAGCCTACGGCGCATTCACGGCGACGCTTCTGCGCGACGACGAGGATTCACAC 145038  
 QY 599 CGGTAGTTTGTAGGACACTTTCAAAATCGGTATGCGGAGTTTAAGGGTAGGGATTAA 658  
 DB 145039 TGATAGTCTTACGACCCCTCAGCAAAATCGGTTTTCGCGGACTGCGGTATGCGG 145098  
 QY 659 TAGGGAAGGGGAAATCGTCTCAGAAATTAACNAGGTGAGACTCCCTCTCAACGTGACT 718  
 DB 145099 CAGGCTGCCCGGAAAGTCAATCGCGGAATGCAAAAATCTGCGCGCTCAATATGAACC 145158  
 QY 719 ACCCTCTCAGGTGATGGCAAAATTTCTCTCACGGAGGGAAGAGAAATTTCTTAATGGAA 778  
 DB 145159 AATTGACCTGACCACTGCCAAATCGCCCTGCGGCACTACGSCATTTATCTCT---GCCA 145215  
 QY 779 AGATACAGGAGTTGTACACAGCGGAGAAAGATGTACGACGAAATGAGAGAAATAGAG 838  
 DB 145216 ACATGACAGCCTGAAACAGGAGGAGGATGTCGCCGAATTTGGGCAAAATATGCC 145275  
 QY 839 GAGTTAGGTTTTTCGAGTAAGGCTTAACCTTCT 872  
 DB 145276 GTCTGAACACTTTCAAGTCAGGCAAACTTCAT 145309

## RESULT 4

AAA81489  
 ID AAA81489 standard; DNA; 837096 BP.

AC AAA81489;

XX 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm\_37 SEQ ID NO:37.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHIR ) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;

PI  
 XX  
 XX  
 DR

WPI; 2000-318079/27.

Isolated nucleotide sequences of *Neisseria meningitidis* which can be used in the diagnosis and treatment of *N. meningitidis* infection and other *Neisseria* infections, for example, *N. gonorrhoea*.

Claim 7; Page 629-865; 1760pp; English.

The present invention describes methods of obtaining immunogenic proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414 represent specifically claimed *Neisseria meningitidis* genomic DNA sequences: AAA81260 to AAA81303 and AAB25620 to AAB25663 represent *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to *Neisseria* bacteria. For example, some of the identified proteins could be components of vaccines against *Meningococcus B*; against all serotypes; and/or against all pathogenic *Neisseriae*. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious *Meningococcus B* vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;

Query Match 8.5%; Score 90.8; DB 21; Length 837096;

Best Local Similarity 47.7%; Pred. No. 7.4e-16;

Matches 331; Conservative 0; Mismatches 357; Indels 6; Gaps 2;

QY 182 AAGCGGTTCTTGGCGATTTTTCGGCGTTAAGGAAGAAATTTAGTTCTCGGTAAACGTT 241  
 DB 209390 AAGCATATAGTTTCGGCTTCGACATCCCGACTGCGCGACATCCGCTGGCAACGGTT 209449  
 QY 242 CGGACGAATCATATACCTCTCAATAGCTATAGTGAACCTTTACATACCGTTTACA 301  
 DB 209450 CGGACGAATCATATACCTCTCAATAGCTATAGTGAACCTTTACATACCGTTTACA 209509  
 QY 302 TACCTGTTCCCACTTCCCATGTACGAGATAAGTCCGAAAGTTCTCGGAAGACCCCTCG 361  
 DB 209510 CAGCGCAACCCAGTTTCTGTCATGTACCGCCACACCGCGCGCTGTACGGCATGTTATG 209569  
 QY 362 TAAAGTTCACTGACGACAAATTTTGATATAGCTTAGAAGAGATTTGAATTAATAG 421  
 DB 209570 TCGGCTTCCACTGACGAGATTTTACCCCACTCCCGCGCTCTCGAAGCCGTCA 209629  
 QY 422 AGAAAGAAACCCGTTCTCGGGTACTTTGTTACCCCAACACCCCGGGAACCTCT 481  
 DB 209630 GGAACACCGCCCTGCCCTGACCTTTATCGCTTACCCCAACACCCCGGCGTATGCT 209689  
 QY 482 TTTCCAGGGAAGATTGA---GGAGATAAGAAACAGGGGTGTTTCTGTGTATAGACG 538  
 DB 209690 TCACGCGTCCCGAAATCGAAGCGCTCATCGAAGCTTCAGACGGCATCGTCTCGTATG 209749  
 QY 539 AAGCCTACTATCATTTACTCCGGAGAAACCTTTCTTGAAGACGCGCTCAAAAGGGAAGATA 598  
 DB 209750 AAGCCTACGGCGCATTCACGGCGACAGCTTCTTCCGCGACGAGGATTCACCAACC 209809  
 QY 599 CGGTAGTTTGTAGGACACTTTCAAAATCGGTATGCGGAGTTTAAAGGTAGGATTAA 658

Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V; Rappuoli R, Pizza M;

Db 209810 TGAATAGCTTTACGACCCCTCAGCAAAATTCGGTTTTCGGGACTCGGTATCGGTTATCGG 209869  
 Qy 659 TAGGGAAGGGGAAATCTCTCAGAAATTAACAAAGGTGAGACTCCCTTCAACGTGACCT 718  
 Db 209870 CAGGCTCCCGAAGTATCGGCGAATGCAAAATTCCTGCGCCCTTACAATATGAACC 209929  
 Qy 719 ACCCTCTCAGGTGATGGCAAAAGTCTCTCAGGAGGGAAGAGAATTCCTAATGGAAA 778  
 Db 209930 AATTGAGCCTGACCACTGCCAAACTCGCCCTGCGGCACTACGGCATTAATCTCT---GCCA 209986  
 Qy 779 AGATACAGGAGGTGTTACACAGAGGAGAAAGGATGTACGACGAAATGAAGAAAATAGAAG 838  
 Db 209987 ACATCGACAGCTGAAAACGACGACGAGGATGTTGCGCGAATGGGCAAAATATGCC 210046  
 Qy 839 GAGTTGAGGTTTTTCCGAGTAAGGCTAACTTCTT 872  
 Db 210047 GTCTGAACACTTTTCAGTCAGGCAAACTTCAT 210080

## RESULT 5

AAV21209/c

ID AAV21209 standard; DNA; 1664976 BP.

XX AC AAV21209;

XX AC AAV21209;

XX DT 10-NOV-1998 (first entry)

XX XX Methanococcus jannaschii circular chromosome.

XX KW Methanococcus jannaschii; methanogenic archaeon; circular chromosome;

XX KW genome; autotrophic; extrachromosomal element; identification; ds.

XX OS Methanococcus jannaschii.

XX PN WO9807830-A2.

XX XX 26-FEB-1998.

XX PF 22-AUG-1997; 97WO-US14900.

XX XX 22-AUG-1996; 96US-0024428.

XX XX (GENO-) INST GENOMIC RES.

XX PA (UNII) UNIV ILLINOIS FOUND.

XX PA (YUJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

XX XX Bult CJ, Smith HO, Venter JC, White OR, Woese CR;

XX XX WPI; 1998-169145/15.

XX XX Complete genome sequence of methano-genic archaeon, Methanococcus

XX PT jannaschii - useful in identification of M. jannaschii genome

XX PT fragment

XX XX Claim 13; Page 152-585; 614pp; English.

XX XX The present sequence represents the complete 1.66-megabase pair genome

XX CC sequence of the Methanococcus jannaschii circular chromosome. The

XX CC present invention describes M. jannaschii open reading frames from the

XX CC genome sequence. The invention also describes a computer based system

XX CC for identifying fragments of the M. jannaschii genome that are

XX CC homologous to target nucleotide sequences, comprising: (a) data storage

XX CC means comprising the nucleotide sequence of the 1664976, 58407 or 16550

XX CC bp sequence (see AAV21209, AAV21210 and AAV21211), or a nucleotide

XX CC sequence at least 99.9% identical to it; (b) search means for comparing a

XX CC target sequence to the nucleotide sequence of the data storage means to

XX CC identify a homologous sequence, and (c) retrieval means for obtaining

XX CC the homologous sequence. The method, which is based on whole genome

XX CC random sequencing of 3 physically distinct elements, a large circular

XX CC chromosome (the 1664976 bp sequence given in AAV21209), a large circular

XX CC extra-chromosomal element (the 58407 bp sequence given in AAV21210), and

XX CC a small circular extra-chromosomal element (the 16550 bp sequence given

CC in AAV21211), can be used in the identification of M. jannaschii genome  
 CC fragment.

XX SQ Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;

Query Match 5.7%; Score 60.2; DB 19; Length 1664976;

Best Local Similarity 47.7%; Pred. No. 2.9e-06;

Matches 479; Conservative 0; Mismatches 493; Indels 33; Gaps 9;

Qy 62 TCAGGCTTCTCTTAACGAATTCCTCAGACTTTCCCGAGGAGATATAAACAAAGGCGCT 121

Db 586070 TAAACTCTGGTCTTAATGAATATCTTGGGACCATCTCCAAATAATAAAGAAAAATTT 586011

Qy 122 TAGAGAATTAATAAAGGTTCCCTTGAACAATATCCAGACCCCGAAGCGAAGAGTTAA 181

Db 586010 TAGATGAAATTGACAAAATTC-----ACCAATATCCAGAGCCAGTAATCCAATTTAA 585957

Qy 182 AAGCGGTTCTTCGGGATTTTTCGGGTTAAGGAGAAAAATTTA---GTTCTCGGTAAAG 238

Db 585956 TGAAGAGTTAAGCAAAATTTTGAATGTTGATGAGGAAAAACATAATGTTGGAGGAGATG 585897

Qy 239 GTTCGACGAACTCATATACTACTCTCAATAGCTATAGGTGAACCTTTACATACCCGTTT 298

Db 585896 GAGCTGATGAGATTATAGACACAATATTTAGAACGTTTGTGATGATGAGATGAGGTTA 585837

Qy 299 ACATACCTGTTCCACCTTCCCATGTACGAGATAAGTCCGAAAGTTCTCGGAAGACCCC 358

Db 585836 TAAATCCAAATTCACAAATTTACCCAATATAGAGTTTCAGCAACAATCCACAATGCTAAAA 585777

Qy 359 TCGTAAAGTTCAACTGGACGAAACTTTGATATAGACTTAGAAGAAGATATGAAATTTAA 418

Db 585776 TAAATATGCTAAATATGATAGGAGAAAGACTTTTAAATTTGAATGTTGAAAGTGTCTTAA 585717

Qy 419 TAGA---GAAAGAAAACCCGTTCTCGGTACTTTTGTCTTACCCAAACACCCCGGAA 475

Db 585716 ATAATATAACAGATAAAACGAAAGTTATTTTCTCTGCACATCCAAATAATCAACAGAA 585657

Qy 476 ACCTCTTTCCAGGGGAAAGATTGA---GGAGATAAGAAACAGGGGTTTCTGTGTAA 532

Db 585656 ATATAATAAGAAATAGATGTAGAGAGGGTTATCAATGAAACACAGACGCTTTAGTTGTTA 585597

Qy 533 TAGCGAAGCCTACTATCATTTACTCCCGAGAAACCTTT-----CTGGAACACCGCTCA 586

Db 585596 TTGACCATGCATACATTGAGTATGCTTAAAGAAATATGATTTGACTCAAAAAGCCCTG 585537

Qy 587 AAAGGAAGATACGGTAGTTTGTAGGACACTTTTCAAAATC---GGTATGCCAGTTTAA 643

Db 585536 AATATGATAATGTTATTTGTTTAAAGACCTTTTCAAAAGTCTTTGTTTAGCAGGAATGA 585477

Qy 644 GGTAGGGATTTTAAATAGGGAAGGGGAAATCGTCTCAGAATTAACAAGGTGAGACTCC 703

Db 585476 GGGTTGGTATGGTGTGCAATAAAAAATCATAGATTATATGATGAGAGTTAAGCCAA 585417

Qy 704 CCTTCAAGCTACCTACCTCTCAGGTGATGCGAAAAGTTCTCTCAGGAGGGAAGAG 763

Db 585416 TATTAGCTTAAACAGGTTAAGTCAAGTTTGTGCCATACTGCATTAGAGAT---AGAG 585360

Qy 764 AATTCCTAATCGAAGAAATACAGAGGTTGTAAACAGCGGAGAAAGGATGTACGACGAAA 823

Db 585359 AATTCTTTGAAAGATGTTTAGAGATGGAATTAAGATAGAGAGATGCTCTACAATGGAT 585300

Qy 824 TGAAGAAAATAGAAGAGTTGAGGTTTTTCCGAGTAAGGCTTAAC---TCTGCTTTTCA 880

Db 585299 TGAAGAGTTTAAAGATATTAAGGTTTATCTTCAAGAGCTTAATATCTATTGTTGAAT 585240

Qy 881 GAACGCTTACCCGCCACGAGGTTTATCAGAGCTACTGAAAAGGATGCTCTCTGCTCA 940

Db 585239 TAAAAACAATGAAACGCAAAAGAAATTTTGTGAGGAACATTTAAAAAGAGGTGTTATTGTTA 585180

Qy 941 GGAAGTATCTTACATCGAAGG---ACTCCAAAAGTCCCTCAGGGTAAGCGTAGGGAAC 997

Db 585179 GAGATTGCACATCTCTTTGATGGTTTATGGGATAATATTGTTAGAGTATCAATAGGACGT 585120





CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;  
Query Match 3.5%; Score 37.6; DB 22; Length 588;  
Best Local Similarity 50.6%; Pred. No. 0.45;  
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 680 CAGAAATTAACAAGGTGAGACTCCCTTCAACGTGACCTACCCCTCTCAGGTGATGCGAA 739  
Db 309 CAGGCTTAGCAATCTGTGGGACAAACATCAAAATGTGCAACTTATGTGTGTAGTTA 368  
QY 740 AAGTTCTCTCAGGAGGAGAGAAATTCCTTAATGGAAGATACAGAGGTTGTAAACAG 799  
Db 369 CAGAAGAGAGATGGTGGGAGGAGGATATTGGAACAAATAATGGCTCAAAACT 428  
QY 800 AGCGAGAAAGGATGTACGACGAATGAAGAAATAGAGAGGTTGAGGTTTCCGAGTA 859  
Db 429 TCCCAATCTGATGAAGACATAAATAAATCAATCAAGAAATTTAAGGAAGTCCAAGTA 488

RESULT 10  
AAK07876  
ID AAK07876 standard; DNA; 588 BP.

XX AC AAK07876;  
XX DT 05-NOV-2001 (first entry)  
XX DE Human brain expressed single exon probe SEQ ID NO: 7867.  
XX DE Human; brain expressed exon; gene expression analysis; probe;  
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX KW epilepsy; cancer; ss.  
XX OS Homo sapiens.  
XX PN WO200157275-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00667.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX PS WPI; 2001-483446/52.  
XX DR Single exon nucleic acid probes for analyzing gene expression in human  
XX PT brains -  
XX PS Example 4; SEQ ID NO: 7867; 650pp + Sequence Listing; English.  
XX CC The present invention provides a number of single exon nucleic acid  
XX CC probes which are derived from genomic sequences expressed in the human  
XX CC brain. They can be used to measure gene expression in brain cell samples,  
XX CC which may enable the diagnosis and improved treatment of nervous system  
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX CC epilepsy and cancers. The present sequence is one of the probes of the  
XX CC invention.  
XX SQ Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;

Query Match 3.5%; Score 37.6; DB 22; Length 588;  
Best Local Similarity 50.6%; Pred. No. 0.45;  
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 680 CAGAAATTAACAAGGTGAGACTCCCTTCAACGTGACCTACCCCTCTCAGGTGATGCGAA 739  
Db 309 CAGGCTTAGCAATCTGTGGGACAAACATCAAAATGTGCAACTTATGTGTGTAGTTA 368  
QY 740 AAGTTCTCTCAGGAGGAGAGAAATTCCTTAATGGAAGATACAGAGGTTGTAAACAG 799  
Db 369 CAGAAGAGAGATGGTGGGAGGAGGATATTGGAACAAATAATGGCTCAAAACT 428  
QY 800 AGCGAGAAAGGATGTACGACGAATGAAGAAATAGAGAGGTTGAGGTTTCCGAGTA 859  
Db 429 TCCCAATCTGATGAAGACATAAATAAATCAATCAAGAAATTTAAGGAAGTCCAAGTA 488

RESULT 11  
AAK33743  
ID AAK33743 standard; DNA; 588 BP.

XX AC AAK33743;  
XX DT 06-NOV-2001 (first entry)  
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 8300.  
XX DE Human; bone marrow expressed exon; gene expression analysis; probe;  
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX OS Homo sapiens.  
XX PN WO200157276-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00668.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX PS WPI; 2001-488900/53.  
XX DR Human genome-derived single exon nucleic acid probes useful for  
XX PT analyzing gene expression in human bone marrow -  
XX PS Example 4; SEQ ID NO: 8300; 658pp + Sequence Listing; English.  
XX CC The present invention provides a number of single exon nucleic acid  
XX CC probes which are derived from genomic sequences expressed in the human  
XX CC bone marrow. They can be used to measure gene expression in bone marrow  
XX CC samples, which may enable the improved diagnosis and treatment of cancers  
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
XX CC the probes of the invention.  
XX SQ Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;

Query Match 3.5%; Score 37.6; DB 22; Length 588;  
Best Local Similarity 50.6%; Pred. No. 0.45;  
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 680 CAGAAATTAACAAGGTGAGACTCCCTTCAACGTGACCTACCCCTCTCAGGTGATGCGAA 739  
Db 309 CAGGCTTAGCAATCTGTGGGACAAACATCAAAATGTGCAACTTATGTGTGTAGTTA 368  
QY 740 AAGTTCTCTCAGGAGGAGAGAAATTCCTTAATGGAAGATACAGAGGTTGTAAACAG 799  
Db 369 CAGAAGAGAGATGGTGGGAGGAGGATATTGGAACAAATAATGGCTCAAAACT 428  
QY 800 AGCGAGAAAGGATGTACGACGAATGAAGAAATAGAGAGGTTGAGGTTTCCGAGTA 859  
Db 429 TCCCAATCTGATGAAGACATAAATAAATCAATCAAGAAATTTAAGGAAGTCCAAGTA 488



Db 309 CAGGCTTAGCAATCTGTGGGACACATCAAACTGTGCCAACTTATGTGTGTGTAGTTA 368  
 Qy 740 AAGTTCTCTCACCAGGAGAGAGAAATTCCTAATGGAAAAGATACAGAGGTTGTACAG 799  
 Db 369 CAGAGAAAGATGTGGGGAGAAAGGATATTGGAAAACAAAATAATGGCTGAAAAC 428  
 Qy 800 AGCAGAAAGATGTACGACGAAATGAAGAAATAGAGAGGTTGAGGTTTTTCCGAGTA 859  
 Db 429 TCCCAATCTGATGAAAGACATTAATATAACATCCCAAGATTTTAAGGAAGTCCAAGTA 488

RESULT 12  
 AAI16517  
 ID AAI16517 standard; DNA; 588 BP.  
 AC AAI16517;  
 XX  
 XX  
 XX 12-OCT-2001 (first entry)  
 DE  
 DE Probe #6450 for gene expression analysis in human cervical cell sample.  
 KW Probe; human; microarray; gene expression; cervical epithelial cell;  
 KW cervical cancer; ss.  
 OS Homo sapiens.  
 XX  
 XX WO200157278-A2.  
 PN  
 PD  
 PD 09-AUG-2001.  
 XX  
 XX 30-JAN-2001; 2001WO-US00670.  
 XX  
 XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-488901/53.  
 XX  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human cervical epithelial cells -  
 PS Claim 25; SEQ ID No 6450; 487pp; English.  
 XX  
 XX The present invention relates to human single exon nucleic acid probes  
 CC (SENP). The present sequence is one such probe. The SENPs are derived  
 CC from human HeLa cells. The SENPs can be used to produce a single exon  
 CC microarray, which can be used for measuring human gene expression in a  
 CC sample derived from human cervical epithelial cells. By measuring gene  
 CC expression, the probes are therefore useful in grading and/or staging  
 CC of diseases of the cervix, notably cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;

Query Match 3.5%; Score 37.6; DB 22; Length 588;  
 Best Local Similarity 50.6%; Pred. No. 0.45;  
 Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy 680 CAGAAATTAACAGGTGACACTCCCTTCAACGTGACCTACCCCTCAGGTGATGGCAA 739  
 Db 309 CAGGCTTAGCAATCTGTGGGACACATCAAAATGTGCCAACTTATGTGTGTGTAGTTA 368  
 Qy 740 AAGTTCTCTCACCAGGAGAGAGAAATTCCTAATGGAAAAGATACAGAGGTTGTACAG 799

Db 369 CAGAGAAAGATGTGGGGAGGAGAAAGATATTGAAAACAAAATAATGGCTGAAAAC 428  
 Qy 800 AGCAGAAAGATGTACGACGAAATGAAGAAATAGAGAGGTTGAGGTTTTTCCGAGTA 859  
 Db 429 TCCCAATCTGATGAAAGACATAAATATAACATCCCAAGATTTTAAGGAAGTCCAAGTA 488

RESULT 13  
 AAI39468  
 ID AAI39468 standard; DNA; 588 BP.  
 AC AAI39468;  
 XX  
 XX 17-OCT-2001 (first entry)  
 DE  
 DE Probe #8154 used to measure gene expression in human placenta sample.  
 KW Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder; ss.  
 OS Homo sapiens.  
 XX  
 XX WO200157272-A2.  
 PN  
 PD  
 PD 09-AUG-2001.  
 XX  
 XX 30-JAN-2001; 2001WO-US00663.  
 XX  
 XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-488997/53.  
 XX  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -  
 PS Claim 25; SEQ ID No 8154; 654pp; English.  
 XX  
 XX The present invention relates to single exon nucleic acid probes (SENP).  
 CC The present sequence is one such probe. The probes are useful for  
 CC producing a microarray for predicting, measuring and displaying gene  
 CC expression in samples derived from human placenta. The probes are useful  
 CC for antenatal diagnosis of human genetic disorders.  
 XX  
 SQ Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;

Query Match 3.5%; Score 37.6; DB 22; Length 588;  
 Best Local Similarity 50.6%; Pred. No. 0.45;  
 Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy 680 CAGAAATTAACAGGTGAGACTCCCTTCAACGTGACCTACCCCTCAGGTGATGGCAA 739  
 Db 309 CAGGCTTAGCAATCTGTGGGACACATCAAAATGTGCCAACTTATGTGTGTGTAGTTA 368  
 Qy 740 AAGTTCTCTCACCAGGAGAGAGAAATTCCTAATGGAAAAGATACAGAGGTTGTACAG 799  
 Db 369 CAGAGAAAGATGTGGGGAGGAGAAAGATATTGAAAACAAAATAATGGCTGAAAAC 428  
 Qy 800 AGCAGAAAGATGTACGACGAAATGAAGAAATAGAGAGGTTGAGGTTTTTCCGAGTA 859  
 Db 429 TCCCAATCTGATGAAAGACATAAATATAACATCCCAAGATTTTAAGGAAGTCCAAGTA 488

RESULT 14

ABS08616

ID ABS08616 standard; DNA; 588 BP.

XX AC

XX ABS08616;

DT 19-AUG-2002 (first entry)

XX

DE Human genome-derived single exon probe from lung SEQ ID NO 8607.

XX

XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;

KW chronic obstructive pulmonary disease; interstitial lung disease;

KW familial idiopathic pulmonary fibrosis; neurofibromatosis;

KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;

KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

KW primary ciliary dyskinesia; pulmonary hypertension;

KW hyaline membrane disease.

XX

OS Homo sapiens.

XX

XX WO200186003-A2.

XX

XX 15-NOV-2001.

XX

XX 30-JAN-2001; 2001WO-US00665.

XX

XX 04-FEB-2000; 2000US-180312P.

PR 26-MAY-2000; 2000US-207456P.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-234687P.

PR 27-SEP-2000; 2000US-236359P.

PR 04-OCT-2000; 2000GB-0024263.

XX

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX

XX WPI; 2002-114183/15.

XX

XX Spatially-addressable set of single exon nucleic acid probes, used to

PT measure gene expression in human lung samples -

XX

XX Claim 1; SEQ ID NO 8607; 634pp; English.

XX

XX The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived

CC from human lung comprising single exon nucleic acid probes having one of

CC 12614 nucleic acid sequences mentioned in the specification, or their

CC complements or the 12387 open reading frames derived from the 12614

CC probes. Also included are a microarray comprising the novel set of

CC probes; the novel set of probes which hybridise at high stringency to a

CC nucleic acid expressed in the human lung; measuring gene expression in a

CC sample derived from human lung, comprising (a) contacting the array with

CC a collection of detectably labeled nucleic acids derived from human lung

CC mRNA, and (b) measuring the label detectably bound to each probe of

CC the array; identifying exons in a eukaryotic genome, comprising

CC (a) algorithmically predicting at least one exon from genomic sequences

CC of the eukaryote; and (b) detecting specific hybridisation of detectably

CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,

CC having a fragment identical to the predicted exon, the probe is included

CC in the above mentioned microarray; assigning exons to a single gene,

CC comprising (a) identifying exons from genomic sequence by the method

CC above and (b) measuring the expression of each of the exons in several

CC tissues and/or cell types using hybridisation to a single exon

CC microarrays having a probe with the exon, where a common pattern of

CC expression of the exons in the tissues and/or cell types indicates that

CC the exons should be assigned to a single gene; a peptide comprising one

CC of 12011 sequences, mentioned in the specification, or encoded by the

CC probes/open reading frames (ORF). The probes are used for gene

CC expression analysis, and for identifying exons in a gene, particularly

CC using human lung derived mRNA and for the study of lung diseases

CC such as asthma, lung cancer, chronic obstructive pulmonary disease

CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary

CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,

CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary

CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,

CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic

CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension

CC and hyaline membrane disease. The present sequence is a single exon

CC probe of the invention.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;

Query Match 3.5%; Score 37.6; DB 24; Length 588;

Best Local Similarity 50.6%; Pred. No. 0.45; Mismatches 0; Gaps 0;

Matches 91; Conservative 0;

QY 680 CAGAAATTAACAAGGTGAGACTCCCTTCAACGTGACCTACCCCTCTCAGGTGATGGCAA 739

DB 309 CAGCCCTTAGCAATCTGTGGACAACATCAATGTGCCAACTTATGTGTGGTAGTTA 368

QY 740 AAGTTCTCTCAGCGAGGGAAGAGAAATTCCTAATGGAAGAATACAGGAGGTTGTACAG 799

DB 369 CAGAAAGAAGAGATGCTGGGGGAGGAAAGATATTGGAACAAATAATGGCTGCAAAACT 428

QY 800 AGCGAGAAGAGATGTACGACGAAATAGAAAGAGTTGAGGTTTTCGAGTA 859

DB 429 TCCCAATCTGATGAAGACATCAATATAACATCCAGAAATTTAAGGAAGTCCAAGTA 488

RESULT 15

ABL21524/C

ID ABL21524 standard; DNA; 9835 BP.

XX

AC ABL21524;

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 16045.

XX

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX

OS Drosophila melanogaster.

XX

XX WO200171042-A2.

XX

PD 27-SEP-2001.

XX

XX 23-MAR-2001; 2001WO-US09231.

XX

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

XX (PEKE ) PE CORP NY.

XX

XX Venter JC, Adams M, Li PWD, Myers EW;

XX

XX WPI; 2001-656860/75.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

XX Claim 1; SEQ ID NO 16045; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 9835 BP; 2538 A; 2196 C; 2372 G; 2729 T; 0 other;

Query Match 3.5%; Score 36.8; DB 23; Length 9835;  
Best Local Similarity 50.0%; Pred. No. 3.4;  
Matches 92; Conservative 0; Mismatches 92; Indels 0; Gaps 0;  
QY 53 CCGCCTCGTCAGGCTTTCCTTAACGAATTCCTCCGAGGAGATAAAAC 112  
Db 3019 CGGCGTTTCCACCAAGGCATTCACCAACGAACCTCAACACATATTTAGCTAAAGAATAGAAC 2960  
QY 113 AAAGGCGCTTAGAAGAAATTAAGAAAGTTCCCTTGAAACAAATACCCAGACCCCGAAGCGA 172  
Db 2959 CAACGAATGAGCAATCTTAAAGAGGTTCCCTGTATAAAGACATCAAGAACACA 2900  
QY 173 AAGAGTTAAAGCGGTTCTTTCGGGATTTTTCGGCGTTAAGGAAGAAATTTAGTTCTCG 232  
Db 2899 TTTAGCTAGAAATTTGAATATATTTATTTAGTATATATTTATTTATTTATTTATTA 2840  
QY 233 GTAA 236  
Db 2839 GTAA 2836

Search completed: April 15, 2003, 14:35:26  
Job time : 1558 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 14:00:46 ; Search time 53 Seconds  
(without alignments)  
5184.298 Million cell updates/sec

Title: US-09-905-173-23

Perfect score: 1065

Sequence: 1 atgatacccccagaggattaa.....cccttcaagctctcttaa 1065

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query No.	Score	Match %	Length	DB ID	Description
1	1065	100.0	1065	1	US-08-599-171A-23	Sequence 23, Appl
2	1065	100.0	1065	2	US-08-646-590B-23	Sequence 23, Appl
3	1065	100.0	1065	3	US-09-069-226-23	Sequence 23, Appl
4	1065	100.0	1065	4	US-09-412-184-23	Sequence 23, Appl
5	61	5.7	7218	1	US-08-232-463-14	Sequence 14, Appl
6	42.2	4.0	1092	2	US-08-646-590B-35	Sequence 35, Appl
7	42.2	4.0	1092	4	US-09-412-184-35	Sequence 35, Appl
8	34.4	3.2	12225	2	US-08-822-445-11	Sequence 11, Appl
9	34.4	3.2	12225	4	US-09-396-540-11	Sequence 11, Appl
10	34.4	3.2	12616	4	US-08-822-445-9	Sequence 9, Appl
11	34.4	3.2	12616	6	US-09-396-540-9	Sequence 9, Appl
12	33.2	3.1	1100	1	US-08-234-939-7	Sequence 7, Appl
13	33.2	3.1	1100	1	US-08-558-865-7	Sequence 7, Appl
14	33.2	3.1	1268	1	US-08-234-939-6	Sequence 6, Appl
15	33.2	3.1	1268	1	US-08-558-865-6	Sequence 6, Appl
16	33.2	3.1	1385	1	US-08-234-939-5	Sequence 5, Appl
17	33.2	3.1	1385	1	US-08-558-865-5	Sequence 5, Appl
18	33.2	3.1	1859	1	US-08-234-939-4	Sequence 4, Appl
19	33.2	3.1	1859	1	US-08-558-865-4	Sequence 4, Appl
20	33.2	3.1	1933	1	US-08-234-939-3	Sequence 3, Appl
21	33.2	3.1	1933	1	US-08-558-865-3	Sequence 3, Appl
22	33.2	3.1	3426	1	US-08-234-939-1	Sequence 1, Appl
23	33.2	3.1	3426	1	US-08-558-865-1	Sequence 1, Appl
24	33.2	3.1	3426	3	US-08-654-025-6	Sequence 3, Appl
25	32.8	3.1	180	4	US-08-894-818B-31	Sequence 31, Appl
26	32.4	3.0	1770	4	US-08-943-731-146	Sequence 146, Appl
27	32.4	3.0	24183	4	US-08-943-731-3	Sequence 3, Appl

C 28 32.2 3.0 477 3 US-09-026-343-12 Sequence 12, Appl  
C 29 32.2 3.0 477 4 US-09-362-871-12 Sequence 12, Appl  
C 30 31.6 3.0 5134 2 US-08-310-912A-157 Sequence 157, Appl  
C 31 31.6 3.0 5134 4 US-09-301-085-157 Sequence 157, Appl  
C 32 31.6 3.0 5134 5 PCT-US95-04589-157 Sequence 157, Appl  
C 33 31.6 3.0 10968 2 US-08-680-327-2 Sequence 2, Appl  
C 34 31.6 3.0 10968 4 US-09-228-246-1 Sequence 1, Appl  
C 35 31.2 2.9 70000 4 US-09-851-896-3 Sequence 3, Appl  
C 36 31 2.9 289 4 US-09-007-005-17 Sequence 17, Appl  
C 37 31 2.9 289 4 US-09-244-796-17 Sequence 17, Appl  
C 38 30.6 2.9 1236 4 US-08-964-652-1 Sequence 1, Appl  
C 39 30.6 2.9 1423 1 US-07-829-954-1 Sequence 1, Appl  
C 40 30.6 2.9 1423 1 US-07-994-423-1 Sequence 1, Appl  
C 41 30.6 2.9 1423 1 US-08-421-891-1 Sequence 1, Appl  
C 42 30.6 2.9 25002 4 US-08-961-527-48 Sequence 48, Appl  
C 43 30.4 2.9 115 1 US-08-488-470A-11 Sequence 11, Appl  
C 44 30.4 2.9 115 1 US-07-946-239-3 Sequence 3, Appl  
C 45 30.4 2.9 115 1 US-08-484-505A-11 Sequence 11, Appl

## ALIGNMENTS

RESULT 1  
US-08-599-171A-23  
; Sequence 23, Application US/08599171A  
; Patent No. 5814473  
; GENERAL INFORMATION:

APPLICANT: WARREN, Patrick V.  
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

ADDRESSEE: CECCHI, STEWART & OLSTEIN

STREET: 6 BECKER FARM ROAD

CITY: ROSELAND

STATE: NEW JERSEY

COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/599, 171A

FILING DATE: Concurrently

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: HERRON, CHARLES J.

REGISTRATION NUMBER: 28,019

REFERENCE/DOCKET NUMBER: 331400-38

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 1065 NUCLEOTIDES

TYPE: NUCLEIC ACID

STRANDEDNESS: SINGLE

TOPOLOGY: LINEAR

MOLECULE TYPE: GENOMIC DNA

US-08-599-171A-23

Query Match 100.0%; Score 1065; DB 1; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATACCCAGAGGATTAGGAAGCTTGAAGCTTACAGACGAGGCTCACTCCCGCTCC 60

```

Db 1 ATGATACCCAGAGGATTAAGAACTTGAAGCTTACAAGCGGAGGTCACTCCCGCCTCC 60
QY 61 GTCAGGCTTTCTCTAAGCAATTCCTTACGACTTCCCGAGGAGATTAACAAAGGCC 120
Db 61 GTCAGGCTTTCTCTAAGCAATTCCTTACGACTTCCCGAGGAGATTAACAAAGGCC 120
QY 121 TTAGAAGAAATTAAGAAAGTTCCCTTGAACAATACCCAGACCCCGAAGCGAAAGAGTTA 180
Db 121 TTAGAAGAAATTAAGAAAGTTCCCTTGAACAATACCCAGACCCCGAAGCGAAAGAGTTA 180
QY 181 AAAGCGGTTCTTCGGGATTTTCGGCGTTAAGGAAGAAATTTAGTTCTCGGTAACGGT 240
Db 181 AAAGCGGTTCTTCGGGATTTTCGGCGTTAAGGAAGAAATTTAGTTCTCGGTAACGGT 240
QY 241 TCGGACGAACTCATATACCTCTCAATAGCTATAGTGAACCTTTACATACCCGTTTAC 300
Db 241 TCGGACGAACTCATATACCTCTCAATAGCTATAGTGAACCTTTACATACCCGTTTAC 300
QY 301 ATACCTGTTCCCACTTCCCATGTACGAGATAAGTGCAGAAAGTTCTCGGAAGACCCCTC 360
Db 301 ATACCTGTTCCCACTTCCCATGTACGAGATAAGTGCAGAAAGTTCTCGGAAGACCCCTC 360
QY 361 GTAAAGGTTCACTGGACGAAACTTTGATATAGCTTGAAGAAGATTTGAATTAATA 420
Db 361 GTAAAGGTTCACTGGACGAAACTTTGATATAGCTTGAAGAAGATTTGAATTAATA 420
QY 421 GAGAAGAAACCCGTTCTCGGTAATTTGATATAGCTTGAAGAAGATTTGAATTAATA 480
Db 421 GAGAAGAAACCCGTTCTCGGTAATTTGATATAGCTTGAAGAAGATTTGAATTAATA 480
QY 481 TTTTCCAGGGAAGATTTGAGGAGATAAGAAACAGGCGTCTTCTGCTGAATAGACGAA 540
Db 481 TTTTCCAGGGAAGATTTGAGGAGATAAGAAACAGGCGTCTTCTGCTGAATAGACGAA 540
QY 541 GCCTACTATCTACTCCGGAAGAACTTTCTGGAAGACCGCTCAAAAGGGAAGATAGC 600
Db 541 GCCTACTATCTACTCCGGAAGAACTTTCTGGAAGACCGCTCAAAAGGGAAGATAGC 600
QY 601 GTAGTTTTCAGGACACTTTCAAAATCGGTATGCGAGTTTAAGGTAGGATTTTAATA 660
Db 601 GTAGTTTTCAGGACACTTTCAAAATCGGTATGCGAGTTTAAGGTAGGATTTTAATA 660
QY 661 GGAAGGGGAAATCGTCTCAGAAATTAACAAGTGAAGTCCCTTCAACGTTACCTAC 720
Db 661 GGAAGGGGAAATCGTCTCAGAAATTAACAAGTGAAGTCCCTTCAACGTTACCTAC 720
QY 721 CCCTCTCAGGTGATGGCAAAAGTTCTCCTCAGGAGGAGAGATTCCTTAATGGAAAG 780
Db 721 CCCTCTCAGGTGATGGCAAAAGTTCTCCTCAGGAGGAGAGATTCCTTAATGGAAAG 780
QY 781 ATACAGGAGTTGTACAGAGCGAAGAGGATGTACGACGAAATGAAGAAATGAAGGA 840
Db 781 ATACAGGAGTTGTACAGAGCGAAGAGGATGTACGACGAAATGAAGAAATGAAGGA 840
QY 841 GTTAGGTTTTCGAGTAAGGCTAATCTTCTGCTTTTCAAGCGCTTACCCGCGCCAC 900
Db 841 GTTAGGTTTTCGAGTAAGGCTAATCTTCTGCTTTTCAAGCGCTTACCCGCGCCAC 900
QY 901 GAGGTTTATCAGGACTACTGAAAGGATGTCCTGTCAGGAGAGATTCCTTACATGGAA 960
Db 901 GAGGTTTATCAGGACTACTGAAAGGATGTCCTGTCAGGAGAGATTCCTTACATGGAA 960
QY 961 GGACTCCAAAGTGCCTCAGGTTAAGCGTAGGGAACCGGGAAGAAACAAAGTTTCTG 1020
Db 961 GGACTCCAAAGTGCCTCAGGTTAAGCGTAGGGAACCGGGAAGAAACAAAGTTTCTG 1020
QY 1021 GAAGCACTGAGGAGAGTATAAATCCCTTCAAGCTCTCTTTAA 1065
Db 1021 GAAGCACTGAGGAGAGTATAAATCCCTTCAAGCTCTCTTTAA 1065

```

RESULT 2

```

US-08-646-590B-23
; Sequence 23, Application US/08646590B
; Patent No. 5962283
; GENERAL INFORMATION:
; APPLICANT: Warren, Patrick V.
; APPLICANT: Swanson, Ronald V.
; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,590B
; FILING DATE: 08-May-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/599,171
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/01094
; FILING DATE: 21-January-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/017001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1065 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1062
US-08-646-590B-23

```

```

Query Match 100.0%; Score 1065; DB 2; Length 1065;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGATACCCAGAGATTAAGAACTTGAAGCTTACAAGCGGAGGTCACTCCCGCCTCC 60
Db 1 ATGATACCCAGAGATTAAGAACTTGAAGCTTACAAGCGGAGGTCACTCCCGCCTCC 60
QY 61 GTCAGGTTTCTCTAAGCAATTCCTTGAACAATACCCAGACCCCGAAGCGAAAGGCC 120
Db 61 GTCAGGTTTCTCTAAGCAATTCCTTGAACAATACCCAGACCCCGAAGCGAAAGGCC 120
QY 121 TTAGAAGAAATTAAGAAAGTTCCCTTGAACAATACCCAGACCCCGAAGCGAAAGAGTTA 180
Db 121 TTAGAAGAAATTAAGAAAGTTCCCTTGAACAATACCCAGACCCCGAAGCGAAAGAGTTA 180
QY 181 AAAGCGGTTCTTCGGGATTTTCGGCGTTAAGGAAGAAATTTAGTTCTCGGTAACGGT 240
Db 181 AAAGCGGTTCTTCGGGATTTTCGGCGTTAAGGAAGAAATTTAGTTCTCGGTAACGGT 240
QY 241 TCGGACGAACTCATATACCTCTCAATAGCTATAGTGAACCTTTACATACCCGTTTAC 300
Db 241 TCGGACGAACTCATATACCTCTCAATAGCTATAGTGAACCTTTACATACCCGTTTAC 300

```



QY 661 GGAAGGGGAAATCGTCTCAGAAATTAACAAGGTGAGACTCCCTTCAACGTCACCTAC 720  
|||||  
Db 661 GGAAGGGGAAATCGTCTCAGAAATTAACAAGGTGAGACTCCCTTCAACGTCACCTAC 720  
QY 721 CCCTCTCAGGTGATGGCAAAAGTTCCTCAGAGGAGGAGAAATTCCTAATGAAAG 780  
|||||  
Db 721 CCCTCTCAGGTGATGGCAAAAGTTCCTCAGAGGAGGAGAAATTCCTAATGAAAG 780  
QY 781 ATACAGGAGGTGTTAACAGAGGAGAAAGATGTACGACGAAATGAAGAAATAGAGGA 840  
|||||  
Db 781 ATACAGGAGGTGTTAACAGAGGAGAAAGATGTACGACGAAATGAAGAAATAGAGGA 840  
QY 841 GTTGAAGTTTCCAGTAAGCTAACTTCTTGTCTTTCAGAACGCTTACCCGCCAC 900  
|||||  
Db 841 GTTGAAGTTTCCAGTAAGCTAACTTCTTGTCTTTCAGAACGCTTACCCGCCAC 900  
QY 901 GAGGTTTATCAGGAGTACTGAAAAGGATGTCCTGTCAGAGAGGTATCTTACATGAA 960  
|||||  
Db 901 GAGGTTTATCAGGAGTACTGAAAAGGATGTCCTGTCAGAGAGGTATCTTACATGAA 960  
QY 961 GGACTCCAAAAGTCTCAGGTAGCGTACGGAACCGGAGAAACCAACAAGTTTCG 1020  
|||||  
Db 961 GGACTCCAAAAGTCTCAGGTAGCGTACGGAACCGGAGAAACCAACAAGTTTCG 1020  
QY 1021 GAAGCACTGGAGGAGATATAAAATCCCTTTTCAAGCTCTCTTTAA 1065  
|||||  
Db 1021 GAAGCACTGGAGGAGATATAAAATCCCTTTTCAAGCTCTCTTTAA 1065

## RESULT 4

US-09-412-184-23  
; Sequence 23, Application US/09412184  
; Patent No. 6268188  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Patrick V.  
; APPLICANT: Swanson, Ronald V.  
; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/412,184  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/646,590  
; FILING DATE: 08-May-1996  
; APPLICATION NUMBER: 08/599,171  
; FILING DATE: 09-FEB-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/01094  
; FILING DATE: 21-January-1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile, Ph.D., Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 09010/017001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1065 base pairs

; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 1...1062  
; US-09-412-184-23  
  
Query Match 100.0%; Score 1065; DB 4; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATATCCCCAGAGGATTAAGGAACTTGAAGCTTACAAGACGAGGTCACTCCCGCCTCC 60  
|||||  
Db 1 ATGATATCCCCAGAGGATTAAGGAACTTGAAGCTTACAAGACGAGGTCACTCCCGCCTCC 60  
QY 61 GTCAAGGTTTCCCTCTACAGAAATTCCTCAGACTTCCCGAGGAGATATAACAAAGGCC 120  
|||||  
Db 61 GTCAAGGTTTCCCTCTACAGAAATTCCTCAGACTTCCCGAGGAGATATAACAAAGGCC 120  
QY 121 TTAGAAGAAATTAAGAAAGTTCCCTTTGAACAAATACCCAGACCCCGAAGGAAAGAGTTA 180  
|||||  
Db 121 TTAGAAGAAATTAAGAAAGTTCCCTTTGAACAAATACCCAGACCCCGAAGGAAAGAGTTA 180  
QY 181 AAGCGGTTCTTGGGATTTTTCGGGTTAAGGAAGAAATTTAGTCTCGTAACGGT 240  
|||||  
Db 181 AAGCGGTTCTTGGGATTTTTCGGGTTAAGGAAGAAATTTAGTCTCGTAACGGT 240  
QY 241 TCGGACGAACATCATATACTACCTCTCAATAGCTATAGTGAACCTTTACATACCCGTTTAC 300  
|||||  
Db 241 TCGGACGAACATCATATACTACCTCTCAATAGCTATAGTGAACCTTTACATACCCGTTTAC 300  
QY 301 ATACCTGTCCCACTTTCCCATGTACGAGATAGTCGGAAGTTCTCGGAAGACCCCTC 360  
|||||  
Db 301 ATACCTGTCCCACTTTCCCATGTACGAGATAGTCGGAAGTTCTCGGAAGACCCCTC 360  
QY 361 GTAAGGTTCAACTGGAGAAACCTTTGATATAGACTTAGAAGAAAGTATTGAATTAATA 420  
|||||  
Db 361 GTAAGGTTCAACTGGAGAAACCTTTGATATAGACTTAGAAGAAAGTATTGAATTAATA 420  
QY 421 GAGAAAGAAACCCGTTCTCGGTTACTTCTGCTTACCCAAACACCCCGGAAACCTC 480  
|||||  
Db 421 GAGAAAGAAACCCGTTCTCGGTTACTTCTGCTTACCCAAACACCCCGGAAACCTC 480  
QY 481 TTTTCCAGGGGAAAGATTTGAGGAGATAGAAACAGGGGTGTTTCTGTGTAATAGACGAA 540  
|||||  
Db 481 TTTTCCAGGGGAAAGATTTGAGGAGATAGAAACAGGGGTGTTTCTGTGTAATAGACGAA 540  
QY 541 GCCTACTATCATTTACTCCGAGAAACCTTTCTGGAAGACCGCTCAAAAGGAGATACG 600  
|||||  
Db 541 GCCTACTATCATTTACTCCGAGAAACCTTTCTGGAAGACCGCTCAAAAGGAGATACG 600  
QY 601 GTAGTTTGTAGGACACTTTCAAAATCGGTATGGGAGTTTAAGGTAGGATTTTAATA 660  
|||||  
Db 601 GTAGTTTGTAGGACACTTTCAAAATCGGTATGGGAGTTTAAGGTAGGATTTTAATA 660  
QY 661 GGAAGGGGAAATCGTCTCAGAAATTAACAAGGTGAGACTCCCTTCAACGTCACCTAC 720  
|||||  
Db 661 GGAAGGGGAAATCGTCTCAGAAATTAACAAGGTGAGACTCCCTTCAACGTCACCTAC 720  
QY 721 CCCTCTCAGGTGATGGCAAAAGTTCCTCAGAGGAGGAGAAATTCCTAATGAAAG 780  
|||||  
Db 721 CCCTCTCAGGTGATGGCAAAAGTTCCTCAGAGGAGGAGAAATTCCTAATGAAAG 780  
QY 781 ATACAGGAGGTGTTAACAGAGGAGAAAGATGTACGACGAAATGAAGAAATAGAGGA 840  
|||||  
Db 781 ATACAGGAGGTGTTAACAGAGGAGAAAGATGTACGACGAAATGAAGAAATAGAGGA 840  
QY 841 GTTGAAGTTTCCAGTAAGCTAACTTCTTGTCTTTCAGAACGCTTACCCGCCAC 900  
|||||  
Db 841 GTTGAAGTTTCCAGTAAGCTAACTTCTTGTCTTTCAGAACGCTTACCCGCCAC 900



```

QY 901 GAGCTTTATCAGGAGTACTGAAAGGATGCTCTGTCAGGAGCGATATCTTACATGGAA 960
Db 901 GAGCTTTATCAGGAGTACTGAAAGGATGCTCTGTCAGGAGCGATATCTTACATGGAA 960
QY 961 GGAATCAAAAGTCCCTCAGGTTAAGCGTAGGGAACCGGAGAAACAAACAAGTTCTG 1020
Db 961 GGAATCAAAAGTCCCTCAGGTTAAGCGTAGGGAACCGGAGAAACAAACAAGTTCTG 1020
QY 1021 GAAGCACTGGAGGAGATATAAATCCCTTTCAAGCTCTCTTTAA 1065
Db 1021 GAAGCACTGGAGGAGATATAAATCCCTTTCAAGCTCTCTTTAA 1065

RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 5.7%; Score 61; DB 1; Length 7218;
Best Local Similarity 1.9%; Pred. No. 2.4e-09;
Matches 7; Conservative 229; Mismatches 139; Indels 0; Gaps 0;

QY 473 GAAACCTCTTTCCAGGGAAGATTGAGGATGAAGAAACAGGCGTCTTCTGTGTA 532
Db 1448 GAAGATTGTTACATCTTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTG 532
QY 533 TAGACGACCGCTACTTACTTACTTCCGCGAGAAACCTTCTGGAAGACCGCTCAAAAGG 592
Db 1388 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 592

```

```

QY 593 AAGATACGCTAGTTTGTAGGACACACTTTCAAAATCGGTATGCGAGTTTAAGGGTAGGA 652
Db 1328 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 652
QY 653 TTTTAATAGGGAAGGGGAAATCGTCTCAGAAATTAACAAGGTGAGACTCCCTTCAACG 712
Db 1268 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 712
QY 713 TGACCTACCCCTCTCAGGTGATGCAAAAGTTCTCTCAGGAGGGAAGAGAAATTCCTAA 772
Db 1208 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 772
QY 773 TCGAAAGATACAGCAGGTTGTACAGAGCGAGAAAGATGTACAGCAATGAAGAAAA 832
Db 1148 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 832
QY 833 TAGAAGGAGTTGAGG 847
Db 1088 RRRRRRRRRRRRR 1074

RESULT 6
US-08-646-590B-35
; Sequence 35, Application US/08646590B
; Patent No. 5962283
; GENERAL INFORMATION:
; APPLICANT: Warren, Patrick V.
; APPLICANT: Swanson, Ronald V.
; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,590B
; FILING DATE: 08-May-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/599,171
; FILING DATE: 09-FEB-1996
; APPLICATION NUMBER: PCT/US97/01094
; FILING DATE: 21-January-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/017001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1092 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1089
US-08-646-590B-35

```





QY 102 GGAGATAAACAAAGGCGCTTAGAAGAAATTAATAAGGTTCCCTTTGAACAAATACCCAGA 161  
Db 8817 GGATTCAAAATCAAGGATATATCTAAATAGCTCCAGATATACCCAGGAGTGCTCT 8876  
QY 162 CCCGAAGCGAAGAGTTAAAGCGGTTCTTCGCGGATTTTTTCGGCGTTAAAGGAAGAAA 221  
Db 8877 CTCCTCAAGGAATGAGAGAAAAGGTGATCCAGCATATTAGAGGAATGTATAAGTAGA 8936  
QY 222 TTTAGTTCTCGGTACGGTTCGGAGCACTCATATACCTTCACCTCAATAGCTATAGGTGA 281  
Db 8937 TTTAGTGCAGCAGACATTTGCAGGAACTTTATTCAGCAGCTGCACACATGATAGAGCAGT 8996  
QY 282 ACTTTACATACCGTTTACATACC 305  
Db 8997 ATGGTATGACCCCATCTACTATCC 9020

## RESULT 11

US-09-396-540-9  
; Sequence 9, Application US/09396540  
; Patent No. 6310182  
; GENERAL INFORMATION:  
; APPLICANT: Kaplan, Jerry  
; APPLICANT: Perou, Charles  
; APPLICANT: Moore, Karen  
; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS  
; AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036/2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/396,540  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/822,445  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-062-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12616 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 190..11592  
; US-09-396-540-9

Query Match 3.2%; Score 34.4; DB 4; Length 12616;  
Best Local Similarity 48.0%; Pred. No. 1.6;  
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 102 GGAGATAAACAAAGGCGCTTAGAAGAAATTAATAAGGTTCCCTTTGAACAAATACCCAGA 161

Db 8817 GGATTCAAAATCAAGGATATATCTAAATAGCTCCAGATATACCCAGGAGTGCTCT 8876  
QY 162 CCCGAAGCGAAGAGTTAAAGCGGTTCTTCGCGGATTTTTTCGGCGTTAAAGGAAGAAA 221  
Db 8877 CTCCTCAAGGAATGAGAGAAAAGGTGATCCAGCATATTAGAGGAATGTATAAGTAGA 8936  
QY 222 TTTAGTTCTCGGTACGGTTCGGAGCACTCATATACCTTCACCTCAATAGCTATAGGTGA 281  
Db 8937 TTTAGTGCAGCAGACATTTGCAGGAACTTTATTCAGCAGCTGCACACATGATAGAGCAGT 8996  
QY 282 ACTTTACATACCGTTTACATACC 305  
Db 8997 ATGGTATGACCCCATCTACTATCC 9020

## RESULT 12

US-08-234-939-7/c  
; Sequence 7, Application US/08234939  
; Patent No. 5459252  
; GENERAL INFORMATION:  
; APPLICANT: Conkling, Mark A.  
; APPLICANT: Yamamoto, Yuri T.  
; TITLE OF INVENTION: Root Specific Gene Promoter  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and Gibson  
; STREET: Post Office Drawer 34009  
; CITY: Charlotte  
; STATE: No. 5459252th Carolina  
; COUNTRY: U.S.A.  
; ZIP: 28234  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/234,939  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/649,564  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sibley, Kenneth D.  
; REGISTRATION NUMBER: 31,665  
; REFERENCE/DOCKET NUMBER: 5051-141  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-881-3140  
; TELEFAX: 919-881-3175  
; TELEX: 575102  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1100 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-234-939-7

Query Match 3.1%; Score 33.2; DB 1; Length 1100;  
Best Local Similarity 51.3%; Pred. No. 0.88;  
Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 267 AATAGCTATAGTGAACCTTTACATACCGTTTACATACCTGTTCACCTTCCACCTTCCCATGTA 326  
Db 331 ATTGATTAAAGATGCTCCACATCCCTTTTACAAACTATTTCTACTCTTTTGTGATTC 272  
QY 327 CGAGATAGTGGGAAGTCTCGGAGAGACCCCTCGTAAAGGTTCAACTGGAGCAAAACTT 386  
Db 271 TTAGATTGTTGTAATATTTCTCTAAAGGAAATAGTTGTCATCAACAGCATTTCTAAAGTA 212



us-09-905-173-23.rni

Wed Apr 16 08:05:51 2003

ADDRESSEE: Gibson  
 STREET: Post Office Drawer 34009  
 CITY: Charlotte  
 STATE: No. 5750386th Carolina  
 COUNTRY: U.S.A.  
 ZIP: 28234  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC Compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/558.865  
 FILING DATE:  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/236.678  
 FILING DATE:  
 CLASSIFICATION: 800  
 APPLICATION NUMBER: US/07/770.082  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Sibley, Kenneth D.  
 REGISTRATION NUMBER: 31,665  
 REFERENCE/DOCKET NUMBER: 5051-166  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-881-3140  
 TELEFAX: 919-881-3175  
 TELEX: 575102  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1268 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-558-865-6

Query Match 3.1%; Score 33.2; DB 1; Length 1268;  
 Best Local Similarity 51.3%; Pred. No. 0.96;  
 Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
 Qy 267 AATAGCTATAGGTGAACCTTTACATACCCGTTTACATACCTGTTCCACCTTTCCCATGTA 326  
 Db 499 ATTTGATTAGATGCTCCTCCACATCCCTTTTACAAACTATTTTCTACTTTTGTGATGTC 440  
 Qy 327 CGAGATAGTCCGAAGCTTCTCGAAGACCCCTCGTAAAGGTTCAACTGGACGAGAAACTT 386  
 Db 439 TTAGATTGTTGTGAATATTTCCCTTAAGGAAATAGTTGCATCAACCCATTTCTAAAGTA 380  
 Qy 387 TGATATAGACTTAGAAGAAAGTATTGAATT 416  
 Db 379 TGCCATAAATTCAAAAATATTTGTGGGTTT 350

Search completed: April 15, 2003, 15:49:10  
 Job time : 100 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 15:15:16 ; Search time 125 seconds  
(without alignments)  
7473.469 Million cell updates/sec

Title: US-09-905-173-23  
Perfect score: 1065  
Sequence: 1 atgataccacagaggattaa.....ccctttcaagctctcttttaa 1065

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438593890 residues  
Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_NEW\_PUB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1065	100.0	1065	9	US-10-060-432-23
2	1065	100.0	1065	10	US-09-905-173-23
3	46.2	4.3	1085	10	US-09-974-300-298
4	42.2	4.0	1092	10	US-09-905-173-35
5	37.8	3.5	198	9	US-10-083-357-579
6	37.6	3.5	588	10	US-09-864-761-6646
7	36.2	3.4	6804	7	US-08-781-986A-365
8	35.8	3.4	560	10	US-09-974-300-4716
9	35.8	3.4	1098	9	US-09-738-626-2298
10	35.2	3.3	1350	9	US-09-971-536-8
11	34.6	3.2	2000	9	US-09-938-842A-3936
12	34.4	3.2	375	10	US-09-864-761-11114
13	34.4	3.2	2270	10	US-09-887-576-455
14	34.4	3.2	12225	10	US-09-927-668-11
15	34.4	3.2	12616	10	US-09-927-668-9
16	34.2	3.2	534	9	US-09-738-626-2011
17	34.2	3.2	1093	9	US-10-001-876-78
18	34.2	3.2	1691139	9	US-10-067-514-1
19	34.2	3.2	3309400	9	US-09-738-626-1

20	33.8	3.2	671	9	US-10-184-644-346
21	33.6	3.2	1851	10	US-09-974-300-2020
22	33.2	3.1	1023	9	US-09-738-626-245
23	33.2	3.1	1480	10	US-09-948-649-1
24	33	3.1	837	10	US-09-974-300-7726
25	33	3.1	948	10	US-09-974-300-7714
26	32.8	3.1	1757	9	US-09-938-842A-4699
27	32.6	3.1	319	10	US-09-974-300-1427
28	32.6	3.1	446	10	US-09-864-761-1815
29	32.6	3.1	53332	9	US-10-224-562-3
30	32.6	3.1	53332	10	US-09-801-861-3
31	32.4	3.0	668	7	US-08-781-986A-309
32	32.4	3.0	1872	10	US-09-815-242-4336
33	32.4	3.0	1884	10	US-09-815-242-8207
34	32.2	3.0	477	12	US-10-028-780-12
35	32	3.0	336	10	US-09-969-708-449
36	32	3.0	336	10	US-09-969-347-312
37	31.8	3.0	472	9	US-10-184-644-98
38	31.6	3.0	161	10	US-09-864-761-23380
39	31.6	3.0	5134	10	US-09-867-852-157
40	31.6	3.0	15772	10	US-09-764-903-66
41	31.6	3.0	32768	10	US-09-070-927A-100
42	31.6	3.0	32816	10	US-09-729-094-3
43	31.4	2.9	242	10	US-09-864-761-21849
44	31.4	2.9	392	9	US-10-184-644-160
45	31.4	2.9	498	10	US-09-864-761-5128

## ALIGNMENTS

RESULT 1  
US-10-060-432-23  
; Sequence 23, Application US/10060432  
; Publication No. US20030040092A1  
; GENERAL INFORMATION:  
; APPLICANT: WARREN, Patrick V.  
; SWANSON, Ronald V.  
; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/060.432  
; FILING DATE: 29-Jan-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/481,733  
; FILING DATE: 11-JAN-2000  
; APPLICATION NUMBER: US/08/599,171  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HERRON, CHARLES J.  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-38  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1065 NUCLEOTIDES  
; TYPE: NUCLEIC ACID

STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: GENOMIC DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 23:  
US-10-060-432-23

Query Match 100.0%; Score 1065; DB 9; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATACCCAGAGGATTAAAGAACTTGAAGCTTACAAGACGAGGTCACTCCGCCCTCC 60  
DB 1 ATGATACCCAGAGGATTAAAGAACTTGAAGCTTACAAGACGAGGTCACTCCGCCCTCC 60

QY 61 GTCAGGCTTCTCTAACAATTCCTTACGACTTCCCGAGGAGATAAACAAGGCC 120  
DB 61 GTCAGGCTTCTCTAACAATTCCTTACGACTTCCCGAGGAGATAAACAAGGCC 120

QY 121 TTAGAAGAATTAATAAGGTTCCCTTGAACAAATACCCAGACCCCGAAGCAAGAGTTA 180  
DB 121 TTAGAAGAATTAATAAGGTTCCCTTGAACAAATACCCAGACCCCGAAGCAAGAGTTA 180

QY 181 AAAGCGGTTCTGCGGATTTTTCGCGGTTAAGGAAGAAATTTAGTCTCGGTAAACGT 240  
DB 181 AAAGCGGTTCTGCGGATTTTTCGCGGTTAAGGAAGAAATTTAGTCTCGGTAAACGT 240

QY 241 TCGGACGAACCTATACACTCTCAATAGCTATAGTGAATTTACATACCGCTTTAC 300  
DB 241 TCGGACGAACCTATACACTCTCAATAGCTATAGTGAATTTACATACCGCTTTAC 300

QY 301 ATACCTGTTCCCACTTCCCATGTACGAGATAAGTGCAGAAAGTTCTCGGAAGACCCCTC 360  
DB 301 ATACCTGTTCCCACTTCCCATGTACGAGATAAGTGCAGAAAGTTCTCGGAAGACCCCTC 360

QY 361 GTAAAGGTTCAACTGACCAAACTTTGATATAGCTTGAAGAAGTATTCGAATTAATA 420  
DB 361 GTAAAGGTTCAACTGACCAAACTTTGATATAGCTTGAAGAAGTATTCGAATTAATA 420

QY 421 GAGAAAGAAACCCGTTCTCGGGTACTTTGCTTACCCAAACACCCACGGGAACCTTC 480  
DB 421 GAGAAAGAAACCCGTTCTCGGGTACTTTGCTTACCCAAACACCCACGGGAACCTTC 480

QY 481 TTTTCAGGGGAAGATTGAGGAGATAAGAAACAGGGGTGTTCTGTCTAATAGACGAA 540  
DB 481 TTTTCAGGGGAAGATTGAGGAGATAAGAAACAGGGGTGTTCTGTCTAATAGACGAA 540

QY 541 GCCTACTATCATTTACTCCGAGAAACCTTTCTGGAAGACGCGCTCAAAAGGAAGATACG 600  
DB 541 GCCTACTATCATTTACTCCGAGAAACCTTTCTGGAAGACGCGCTCAAAAGGAAGATACG 600

QY 601 GTAGTTTGAAGACACTTTCACAAATCGGTATGCGGAGTTTAAAGGTAGGATTTAATA 660  
DB 601 GTAGTTTGAAGACACTTTCACAAATCGGTATGCGGAGTTTAAAGGTAGGATTTAATA 660

QY 661 GGGAGGGGAATCGTCTCAGAAATTAACAAGTGAGACTCCCTTCAACGTGACCTAC 720  
DB 661 GGGAGGGGAATCGTCTCAGAAATTAACAAGTGAGACTCCCTTCAACGTGACCTAC 720

QY 721 CCCTCTCAGGTGATGGCAAAATTCCTCAGGAGGAGAGAAATTCCTAATGGAAG 780  
DB 721 CCCTCTCAGGTGATGGCAAAATTCCTCAGGAGGAGAGAAATTCCTAATGGAAG 780

QY 781 ATACAGAGGTTGTAACAGAGCGAGAAAGGATGTACACGAAATGAAGAAATAGAAGA 840  
DB 781 ATACAGAGGTTGTAACAGAGCGAGAAAGGATGTACACGAAATGAAGAAATAGAAGA 840

QY 841 GTTCAGGTTTTCGAGTAAAGGCTTAACCTTCTGCTTTTCAAGACGCTTACCCGCCAC 900  
DB 841 GTTCAGGTTTTCGAGTAAAGGCTTAACCTTCTGCTTTTCAAGACGCTTACCCGCCAC 900

QY 901 GAGGTTTATCAGGAGCTACTGAAAAGGATGTCTCTCGTCAGGAACGTTTACATGGAA 960  
DB 901 GAGGTTTATCAGGAGCTACTGAAAAGGATGTCTCTCGTCAGGAACGTTTACATGGAA 960

QY 961 GGACTCCAAAAGTCCCTCAGGTAAGCGTAGGAAACCGGAAGAAACAAAGTTTCTG 1020  
DB 961 GGACTCCAAAAGTCCCTCAGGTAAGCGTAGGAAACCGGAAGAAACAAAGTTTCTG 1020

QY 1021 GAACACTGAGGAGAGATATAAATCCCTTCAAGCTCTCTTTAA 1065  
DB 1021 GAACACTGAGGAGAGATATAAATCCCTTCAAGCTCTCTTTAA 1065

## RESULT 2

US-09-905-173-23  
; Sequence 23, Application US/0905173  
; Patent No. US20020132295A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: SHORT, Jay M.  
; APPLICANT: WARREN, Patrick V.  
; APPLICANT: SWANSON, Ronald V.  
; APPLICANT: MATHUR, Eric J.  
; TITLE OF INVENTION: ENZYMES HAVING TRANSAMINASE AND AMINOTRANSFERASE ACTIVITY AND  
; FILE REFERENCE: DIVER1240-7  
; CURRENT APPLICATION NUMBER: US/09/905,173  
; PRIOR FILING DATE: 2001-07-12  
; PRIOR APPLICATION NUMBER: US 09/412,184  
; PRIOR FILING DATE: 1999-10-04  
; PRIOR APPLICATION NUMBER: US 09/389,537  
; PRIOR FILING DATE: 1999-09-02  
; PRIOR APPLICATION NUMBER: US 08/646,590  
; PRIOR FILING DATE: 1996-05-08  
; PRIOR APPLICATION NUMBER: US 08/599,171  
; PRIOR FILING DATE: 1996-02-09  
; PRIOR APPLICATION NUMBER: US 09/481,733  
; PRIOR FILING DATE: 2000-01-11  
; PRIOR APPLICATION NUMBER: US 09/069,226  
; PRIOR FILING DATE: 1998-04-27  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 23  
; LENGTH: 1065  
; TYPE: DNA  
; ORGANISM: Aquifex  
US-09-905-173-23

Query Match 100.0%; Score 1065; DB 10; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATACCCAGAGGATTAAAGAACTTGAAGCTTACAAGACGAGGTCACTCCGCCCTCC 60  
DB 1 ATGATACCCAGAGGATTAAAGAACTTGAAGCTTACAAGACGAGGTCACTCCGCCCTCC 60

QY 61 GTCAGGCTTCTCTAACAATTCCTTACGACTTCCCGAGGAGATAAACAAGGCC 120  
DB 61 GTCAGGCTTCTCTAACAATTCCTTACGACTTCCCGAGGAGATAAACAAGGCC 120

QY 121 TTAGAAGAATTAATAAGGTTCCCTTGAACAAATACCCAGACCCCGAAGCAAGAGTTA 180  
DB 121 TTAGAAGAATTAATAAGGTTCCCTTGAACAAATACCCAGACCCCGAAGCAAGAGTTA 180

QY 181 AAAGCGGTTCTGCGGATTTTTCGCGGTTAAGGAAGAAATTTAGTCTCGGTAAACGT 240  
DB 181 AAAGCGGTTCTGCGGATTTTTCGCGGTTAAGGAAGAAATTTAGTCTCGGTAAACGT 240

QY 241 TCGGACGAACCTATACACTCTCAATAGCTATAGTGAATTTACATACCGCTTTAC 300  
DB 241 TCGGACGAACCTATACACTCTCAATAGCTATAGTGAATTTACATACCGCTTTAC 300

QY 301 ATACCTGTTCCCACTTCCCATGTACGAGATAAGTGCAGAAAGTTCTCGGAAGACCCCTC 360  
DB 301 ATACCTGTTCCCACTTCCCATGTACGAGATAAGTGCAGAAAGTTCTCGGAAGACCCCTC 360



Qy	361	GTAAAGGTTCAACTGGACGAAAACTTTGATATAGACTTGAAGAAGTATTGAATTAATA	420
Db	361	GTAAAGGTTCAACTGGACGAAAACTTTGATATAGACTTGAAGAAGTATTGAATTAATA	420
Qy	421	GAGAAAGAAAAACCGGTTCTCGGGTACTTTGCTTACCCAAACAACCCACGCGGAACCTC	480
Db	421	GAGAAAGAAAAACCGGTTCTCGGGTACTTTGCTTACCCAAACAACCCACGCGGAACCTC	480
Qy	481	TTTTTCCAGGGGAAAGATTGAGGAGATAAGAAACAGGGGTGTTTTCTGTGTATAGACGAA	540
Db	481	TTTTTCCAGGGGAAAGATTGAGGAGATAAGAAACAGGGGTGTTTTCTGTGTATAGACGAA	540
Qy	541	GCCTACTATCATTTACTCCGGAGAAAACTTTCTGGGAAGACGCGCTCAAAAGGCGAATACG	600
Db	541	GCCTACTATCATTTACTCCGGAGAAAACTTTCTGGGAAGACGCGCTCAAAAGGCGAATACG	600
Qy	601	GTAGTTTTTGAGGACACTTTTCAAAAATCGGTATGGCGAGTTTAAGGTAGGGATTTTAATA	660
Db	601	GTAGTTTTTGAGGACACTTTTCAAAAATCGGTATGGCGAGTTTAAGGTAGGGATTTTAATA	660
Qy	661	GGGAAGGGGAAATCGTCTCAGAAAAATTACAGGTTGAGACTCCCCTTCAACGTGACCTAC	720
Db	661	GGGAAGGGGAAATCGTCTCAGAAAAATTACAGGTTGAGACTCCCCTTCAACGTGACCTAC	720
Qy	721	CCCTCTCAGGTGATGGCAAAAGTTCTCCTCAGGAGGGAAGAGAATTCCTAATGAAAAG	780
Db	721	CCCTCTCAGGTGATGGCAAAAGTTCTCCTCAGGAGGGAAGAGAATTCCTAATGAAAAG	780
Qy	781	ATACAGGAGTTGTACAGAGCGAGAAAGATGTACGACGAAATGAAGAAATAGAAGGA	840
Db	781	ATACAGGAGTTGTACAGAGCGAGAAAGATGTACGACGAAATGAAGAAATAGAAGGA	840
Qy	841	GTTGAGGTTTTTCCGAGTAGGCTAACTTCTTGCTTTTCAGAACGCTTTACCCGCCAC	900
Db	841	GTTGAGGTTTTTCCGAGTAGGCTAACTTCTTGCTTTTCAGAACGCTTTACCCGCCAC	900
Qy	901	GAGGTTTATCAGGAGCTACTGAAAAGGATGTCCTCGTCAGGAACGTATCTTACATGGA	960
Db	901	GAGGTTTATCAGGAGCTACTGAAAAGGATGTCCTCGTCAGGAACGTATCTTACATGGA	960
Qy	961	GGACTCCAAAAGTGCTTCAGGTTAAGCGTAGGGAACCGGAAGAAACAACAGTTTCTG	1020
Db	961	GGACTCCAAAAGTGCTTCAGGTTAAGCGTAGGGAACCGGAAGAAACAACAGTTTCTG	1020
Qy	1021	GAGACTGGAGGAGAGTATAAAATCCCTTTTCAAGCTCTCTTTAA	1065
Db	1021	GAGACTGGAGGAGAGTATAAAATCCCTTTTCAAGCTCTCTTTAA	1065

### RESULT 3

US-09-974-300-298  
; Sequence 298, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:

```

/ / Generated information.
/ /
/ / APPLICANT: Berka, Randy M.
/ /
/ / APPLICANT: Clausen, Ib Groth
/ /
/ / TITLE OF INVENTION: Methods For Monitoring Multiple Gene
/ /
/ / TITLE OF INVENTION: Expression
/ /
/ / FILE REFERENCE: 10085.500-US
/ /
/ / CURRENT APPLICATION NUMBER: US/09/974,300
/ /
/ / CURRENT FILING DATE: 2001-10-05
/ /
/ / PRIOR APPLICATION NUMBER: 09/680,598
/ /
/ / PRIOR FILING DATE: 2000-10-06
/ /
/ / PRIOR APPLICATION NUMBER: 60/279,526
/ /
/ / PRIOR FILING DATE: 2001-03-27
/ /
/ / NUMBER OF SEQ ID NOS: 8481
/ /
/ / SOFTWARE: FastSeq for Windows Version 4.0
/ /
/ / SEQ ID NO 298

```

```

; SEQ ID NO 298
; LENGTH: 1085
; TYPE: DNA
; ORGANISM: Bac
US-09-974-300-298

```

```

Query Match      4.3%  Score 46.2;  DB 10;  Length 1085;
Best Local Similarity 47.4%;  Pred. No. 0.00077;
Matches 138;  Conservative 0;  Mismatches 153;  Indels 0;  Gaps 0;

QY 154 TACCACGACC CGCGAAGCGAAGAGTTAAACCGGTTCTTTCGGGATTTTTCGGCGGTTAAG 213
Db 181 TATCCTGACGGCTATAGTCGGCATTTGAGACAAGCTTGTCTCCCATCTCGGCGTAAT 240

QY 214 GAAGAAAAATTTAGTTCGTGGTATCGGGTTCGGACGAACCTCATATACTACCTCTCAATAGCT 273
Db 241 GAAACGAACATCATCTCTGGGAAGCGGAACGACGAAGTCATTCAAAATATTATTCGCCGTCT 300

QY 274 ATAGGTGAACCTTTACATACCCGTTTACATACCTCTGTCCACCTTTTCCCATCTACGAGATA 333
Db 301 TTGTTGGACCCCTGCCCTCAACACACTGTATGCGGAATCGACATTTTCACAATACAAGCAC 360

QY 334 AGTCGGAAGATTCGCGNAGACCCCTCGTAAAGGTTTCAACTGGACGAAAACCTTTGATATA 393
Db 361 AACCGTGTGTCGGAAGGCGGGAAGTCAGAGAATCGGCGCTGCTCGAAAACGGCTGCCAC 420

QY 394 GACTTAGAAAAGATGATTGAAATTAATAGAGAAAAGAAAACCCGTTCTTCGGG 444
Db 421 GACCTTGATCAATGCTGAAGCGCATCGATGAACAGACGAGGTTGTCTGG 471

RESULT 4
US-09-905-173-35
; Sequence 35, Application US/09905173
; Patent No. US20020132295A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: SHORT. Jav M.

```

## RESULT 4

US-09-905-173-35

```

? Sequence 35, Application US/09905173
? Patent No. US20020132295A1
? GENERAL INFORMATION:
? APPLICANT: DIVERSA CORPORATION
? APPLICANT: SHORT, Jay M.
? APPLICANT: WARREN, Patrick V.
? APPLICANT: SWANSON, Ronald V.
? APPLICANT: MATHUR, Eric J.
? TITLE OF INVENTION: ENZYMES HAVING TRANSAMINASE AND AMINOTRANSFERASE ACTIVITY AND
? FILE OF INVENTION: USE THEREOF
? FILE REFERENCE: DIVER1240-7
? CURRENT APPLICATION NUMBER: US/09/905,173
? CURRENT FILING DATE: 2001-07-12
? PRIOR APPLICATION NUMBER: US 09/412,184
? PRIOR FILING DATE: 1999-10-04
? PRIOR APPLICATION NUMBER: US 09/389,537
? PRIOR FILING DATE: 1999-09-02
? PRIOR APPLICATION NUMBER: US 08/646,590
? PRIOR FILING DATE: 1996-05-08
? PRIOR APPLICATION NUMBER: US 08/599,171
? PRIOR FILING DATE: 1996-02-09
? PRIOR APPLICATION NUMBER: US 09/481,733
? PRIOR FILING DATE: 2000-01-11
? PRIOR APPLICATION NUMBER: US 09/069,226
? PRIOR FILING DATE: 1998-04-27
? NUMBER OF SEQ ID NOS: 40
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 35
? LENGTH: 1092
? TYPE: DNA
? ORGANISM: Ammonifex degensii
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (986)..(987)
? OTHER INFORMATION: n is any nucleotide
? US-09-905-173-35

```

Query Match	4.0%	Score 42.2	DB 10	Length 1092
Best Local Similarity	48.2%	Pred. No. 0.013		
Matches 119	Conservative	0	Mismatches 128	Indels 0
				Gaps 0

[illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

STREET: 9410 Key West Ave







Best Local Similarity 48.0%; Pred. No. 13;		Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;	
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;			
QY	102	GGAGATAAAACAAAGGCTTTAGAGAAATTAATAAGGTTCCCTTGAACAAATACCCAGA	161
Db	8430	GGATTCAAATCAAGGATATATCTAAATAGCTCCAGATATACCCAGGCGTGTCTCT	8489
QY	162	CCCCGAAGCGAAGAGTTAAAGCGGTTCTTCGGGATTTTTCGGCGTTAAGGAAGAAA	221
Db	8490	CTCCCAAGGAATGAGAGAAAAAGGTGATCCAGCATATTAGAGGAATGTATAAGTAGA	8549
QY	222	TTTAGTTCTCGTAAAGGTTCCGAGCACTCATATACCTCTCAATAGCTATAGGTGA	281
Db	8550	TTTGAGTCCGACGACATTTGCGAGAACTTATTCAGCAGCTGACACATGATAGAGCAGT	8609
QY	282	ACTTTACATACCCGTTTACATACC	305
Db	8610	ATGGTATGACCCCATCTACTATCC	8633

RESULT 15  
US-09-927-668-9  
; Sequence 9, Application US/09927668  
; Patent No. US20020115144A1  
; GENERAL INFORMATION:  
; APPLICANT: Kaplan, Jerry  
; Moore, Charles  
; Perou, Karen  
; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS  
; AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036/2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/927,668  
; FILING DATE: 10-Aug-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/396,540  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-062-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12616 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 190..11592  
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
US-09-927-668-9

Query Match 3.2%; Score 34.4; DB 10; Length 12616;  
Best Local Similarity 48.0%; Pred. No. 14;  
US-09-927-668-9

Best Local Similarity 48.0%; Pred. No. 13;		Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;	
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;			
QY	102	GGAGATAAAACAAAGGCTTTAGAGAAATTAATAAGGTTCCCTTGAACAAATACCCAGA	161
Db	8430	GGATTCAAATCAAGGATATATCTAAATAGCTCCAGATATACCCAGGCGTGTCTCT	8489
QY	162	CCCCGAAGCGAAGAGTTAAAGCGGTTCTTCGGGATTTTTCGGCGTTAAGGAAGAAA	221
Db	8490	CTCCCAAGGAATGAGAGAAAAAGGTGATCCAGCATATTAGAGGAATGTATAAGTAGA	8549
QY	222	TTTAGTTCTCGTAAAGGTTCCGAGCACTCATATACCTCTCAATAGCTATAGGTGA	281
Db	8550	TTTGAGTCCGACGACATTTGCGAGAACTTATTCAGCAGCTGACACATGATAGAGCAGT	8609
QY	282	ACTTTACATACCCGTTTACATACC	305
Db	8610	ATGGTATGACCCCATCTACTATCC	8633

Search completed: April 15, 2003, 16:58:26  
Job time : 144 secs

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 13:59:36 ; Search time 2156 seconds

(without alignments)  
8000.090 Million cell updates/sec

Title: US-09-905-173-23

Perfect score: 1065

Sequence: 1 atgatacccccaggaggattaa.....ccctttcaagctctcttttaa 1065

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_esti:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65.8	6.2	573	10	BE432349
2	65	6.1	587	10	AW034522
3	62.6	5.9	819	12	BG350830
4	58.8	5.5	762	13	BM412207
5	58.4	5.5	631	13	BJ463860
6	57.2	5.4	778	12	BG887294
					BE432349 EST398878
					AW034522 EST278138
					BG350830 098611 Ma
					BM412207 EST386534
					BJ463860 BJ463860
					BG887294 EST513145

7	53.4	5.0	811	12	BG598806
8	53.2	5.0	622	13	BM322300
9	52.4	4.9	639	14	BM812973
10	51.6	4.8	555	13	BJ280409
11	50.2	4.7	405	13	BM173372
12	47.8	4.5	511	13	BI1717525
13	47.2	4.4	667	14	B0115402
14	47.2	4.4	1757	11	AV105176
15	46.2	4.3	662	13	BM003279
16	45.6	4.3	997	17	CNS005TE
17	44.6	4.2	940	17	CNS06YZU
18	44.2	4.2	925	17	CNS00IBN
19	43.6	4.1	299	9	AI423192
20	43	4.0	313	10	BE341733
21	42.6	4.0	515	13	BI934083
22	42.6	4.0	659	13	BI934806
23	41.6	3.9	540	10	AW065601
24	41.6	3.9	1101	17	CNS017KE
25	40.6	3.8	322	9	AI276240
26	40.2	3.8	588	13	BI994014
27	40	3.8	532	13	BJ128028
28	40	3.8	556	13	BJ108720
29	40	3.8	625	13	BJ104535
30	40	3.8	658	13	BJ120960
31	39.8	3.7	1101	17	CNS0106X
32	39.4	3.7	860	17	A2678707
33	39.4	3.7	889	17	A2677419
34	39.4	3.7	901	17	BI161744
35	39.4	3.7	902	17	A2543474
36	39.4	3.7	920	17	BI151132
37	39.2	3.7	418	12	BG509907
38	39.2	3.7	565	14	BQ134497
39	38.4	3.6	612	13	BJ117344
40	38.4	3.6	631	13	BJ114619
41	38.4	3.6	1101	17	CNS00G3K
42	37.8	3.5	583	14	BQ123162
43	37.8	3.5	873	17	CNS018MF
44	37.8	3.5	898	17	BI133728
45	37.6	3.5	1101	17	CNS00BJ6

#### ALIGNMENTS

RESULT 1	BE432349	EST398878	tomato breaker fruit, TIGR Lycopersicon esculentum cDNA	573 bp	linear	EST 18-MAY-2001
LOCUS	BE432349	clone cLEG7J1, mRNA sequence.				
DEFINITION	BE432349	EST				
ACCESSION	BE432349	GI:9430192				
VERSION	BE432349.1					
KEYWORDS	EST					
SOURCE	tomato.					
ORGANISM	Lycopersicon esculentum					
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;					
AUTHORS	Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.					
TITLE	Generation of ESTs from tomato fruit tissue, breaker stage					
JOURNAL	Unpublished (2000)					
COMMENT	Contact: CUGI					
	Clemson University Genomics Institute					
	Clemson University					
	100 Jordan Hall, Clemson, SC 29634, USA					
	Email: <a href="http://www.genome.clemson.edu/orders/index.html">http://www.genome.clemson.edu/orders/index.html</a>					
FEATURES	5 prime sequence.					
	Location/Qualifiers					
	1. .573					





# AUTHORS TITLE JOURNAL COMMENT

Nielsen, K.L., Crookshanks, M., Emmersen, J. and Welinder, K.G.  
EST-sequencing of mature potato tuber (Var. Kuras)  
Unpublished (2000)  
Contact: Karen G. Welinder  
Institut for bioteknologi  
Aalborg Universitet  
Sohnsgaardsholmsvej 49, 9000 Aalborg, Denmark  
Tel: +45 96358467  
Fax: +45 98141808  
Email: kqw@bio.auc.dk  
Sequenced from the 5' end.  
High quality sequence stop: 819  
POLYA-No.

## FEATURES source

Location/Qualifiers  
1. .819  
/organism="Solanum tuberosum"  
/cultivar="Field grown Kuras"  
/db\_xref="taxon:4113"  
/clone\_lib="Mature tuber lambda ZAP"  
/tissue\_type="Tuber"  
/note="Vector: Lambda ZAP"  
227 a 155 c 194 g 243 t

Query Match 5.9%; Score 62.6; DB 12; Length 819;  
Best Local Similarity 47.0%; Pred. No. 2.6e-06;  
Matches 194; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 149 ACAATACCCAGACCCGACGAAAGAGTTAAAGCGGTCTTCGCGGATTTTCGGCG 208  
Db 393 ATATTATCTCGATCCTGGAAGCGGTAGCTGCGTGCAGCTCTTGCTGAAGATTCTCGCC 452  
QY 209 TTAAGGAAGAAATTTAGTTCTCGGTAAACGGTTCGGACGACTCATATACCTCTCA 268  
Db 453 TTGAATCTGAGTATATCTTCAGCGTCCGTCGAGATGAATCATTTGATATGA 512  
QY 269 TAGCTATAGGTGAACCTTTACATACCGGTTTACATACCTGTTCCACCTTTCCCATGACG 328  
Db 513 GATCGGTATGGATCTCTGTTGACAAGATTGTTGACTGCCACCCACTTTCACAATGATG 572  
QY 329 AGATAAGTGGCAAGTTCTCGGAAGACCCCTCGTAAAGGTTCACTGACGACAACTTTG 388  
Db 573 AATTGATGACGCTGTAATGGACAGGTGTCATCAAGGTGCTAGGAACCCAGACTTTA 632  
QY 389 ATATAGACTTAGAAGAGTATTGAATTAATAGAGAAGAAACCCGTTCTCGGGTACT 448  
Db 633 GCCTGGATGTAGAAGGATGTCGCAAGTGTGAACCTTGAGAACCCAGCTATTTTC 692  
QY 449 TTGCTTACCCAAACACCCGACGAAACCTCTTTTCCAGGGGAAAGATTGAGGAGATA 508  
Db 693 TGACATCACCTAATATCTCTGATGGAGTATTAATTGATGATGAACCTTATTGAAAATAC 752  
QY 509 GAAACAGGGGTGTTTCTGTGTTATAGACGACGCTACTATCATCTACTCTCCGGA 561  
Db 753 TTCACCTGCCCATATTGGTAAATATGGATGAAGCATATGTTGAGTTTCTCGGA 805

RESULT 4  
BM412207 762 bp mRNA linear EST 22-JAN-2002  
LOCUS  
DEFINITION  
EST586534 tomato breaker fruit Lycopersicon esculentum cDNA clone  
CLEG59E12 5' end, mRNA sequence.  
BM412207

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
tomato.  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 762)

REFERENCE  
AUTHORS  
Alcala, J., Vrebalov, J., White, R., Vision, T., Karamycheva, S.A., Tsai

# TITLE JOURNAL COMMENT

J., Bougri, O., Kirkness, E., Utterback, T., Van Aken, S., Ronning  
C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.  
Generation of ESTs from tomato fruit tissue, breaker stage (2002)  
Unpublished (2002)  
Contact: CUGI  
Clemson University Genomics Institute  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
This clone is available through the Clemson University Genomics  
Institute  
Seq primer: T3.

## FEATURES source

Location/Qualifiers  
1. .762  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CLEG59E12"  
/clone\_lib="tomato breaker fruit"  
/tissue\_type="Pericarp"  
/dev\_stage="breaker"  
/lab\_host="SOLR"  
/note="Vector: pBluescriptSMCUadapt; Site\_1: EcoRI;  
Site\_2: XhoI; supplier: Boyce Thompson Institute;  
sequencing: The Institute for Genomic Research. Fruit  
were harvested at the breaker stage (first sign of  
lycopene accumulation on the blossom end of fruit). Fruit  
were cut in half and the seeds and locules were discarded  
prior to freezing the pericarp."  
207 a 155 c 178 g 222 t

BASE COUNT 207 a 155 c 178 g 222 t  
ORIGIN  
Query Match 5.5%; Score 58.8; DB 13; Length 762;  
Best Local Similarity 48.8%; Pred. No. 2.9e-05;  
Matches 159; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 149 ACAATACCCAGACCCGACGAAAGAGTTAAAGCGGTCTTCGCGGATTTTTCGGCG 208  
Db 395 ACATTTATCTGTATCTGAAAGCCGACGCTCGCTGAGATTCTGGCC 454  
QY 209 TTAAGGAAGAAATTTAGTTCTCGGTAAACGGTTCGGACGAACTCATATACCTCTCAA 268  
Db 455 TTGAATCTGAGTATATCTTTCAGGGTTCGAGTGAACCTCATTTGATAATGA 514  
QY 269 TACCTATAGTGAACCTTACATACCGTTTACATACCTGTTCCACCTTTCCCATGTAG 328  
Db 515 GATGCATATTGATCTCTGTTGACAAAGATTGTTGACTGCCACCCCTTTTCAATG 574  
QY 329 AGATAAGTGGCAAGTTCTCGGAAGACCCCTCGTAAAGGTTCAACTGGACGAAACTTTG 388  
Db 575 AATTTGATGACCTGTAATGGAGCAGGTGTCATCAAGTGCCTAGGAACCCAGACTTTA 634  
QY 389 ATATAGACTTAGAAGAGTATTGAATTAATAGAGAAGAAACCCGTTCTCGGGTACT 448  
Db 635 GCCTGACGCTTGAACGATTACCGAAGTGTGGAACGTGAGAAACAAAGTCATATTTCT 694  
QY 449 TTGCTTACCAACACACCCACGGA 474  
Db 695 GACATCACCTAATAATCTGATGGGA 720

RESULT 5  
BJ463860

LOCUS  
DEFINITION

BJ463860 631 bp mRNA linear EST 23-MAY-2002  
K. Sato unpublished cDNA library, cv. Haruna Nijo  
germination shoots Hordeum vulgare subsp. vulgare cDNA clone  
bags30m21 5', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BJ463860.1 GI:21142367  
Hordeum vulgare subsp. vulgare.  
Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;





Db 352 GTACTTGAACCCAGGTGACAAAATTGTAGATTGCCCTCCAACCTTTACCATGTATGAGTTT 411

Db 352 GTACTTGAACCAAGGTGACAAAATGTAGATGGCCCTCCAACCTTTTACCATGTAATGAGTTT 411  
1

[illegible]



Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP + sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT 93 a 164 c 170 g 84 t

Query Match 4.5%; Score 47.8; DB 13; Length 511;  
Best Local Similarity 44.6%; Pred. No. 0.028;  
Matches 187; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

QY 149 ACAATATCCAGACCCGAGCAAGAGATTAAGCGGTTCTTCGGGATTTTTCGGCG 208  
Db 77 ACATCTACCCGAGCCGAGACCCGCGCTTGGGAGAGCGTTGGGAGATGCACAACA 136  
QY 209 TTAAGGAGAAATTTAGTTCTCGGTAACGGTTTCGACGAGCTCATATACCTCTCAA 268  
Db 137 TCCGATGGAGAACCTGCTGGTGGGTCGCGCGGAGAGCTGATCGACCTGCTATGC 196  
QY 269 TAGCTATAGTGAACTTTACATACCGGTTTACATACCTGTTCCACCTTTTCCCATGTAGC 328  
Db 197 GCTGTGCTGGAGCCGCGTGACAAGATCGTGACTGCCCGCCACCTTCACCATGTAGC 256  
QY 329 AGATAAGTGGAAAGTTCTCGGAAGACCCCTCGTAAGGTTCACTGGGAGAACTTTG 388  
Db 257 TGTGTGATGCGGCGTGAACGACGCGCGGTGTGACGGTGGCGCGCTGGAGGCTCC 316  
QY 389 ATATAGACTTTAGAAAGATTTGAATTAATAGAGAAAGAAACCCGTTCTCGGTTACT 448  
Db 317 GCATCGACGTGGAGCGCTCAAGCGTGGTGGTGACACAGCTAAGGTGTCTCC 376  
QY 449 TTGCTTTACCAACAACCCGAGAACTCTTTTCCAGGGGAAGATTTGAGGAGATTA 508  
Db 377 TGACCAAGCCCAACACCCGAGGCTCCATGATCAGCGAGCGGCGACCTGGTCTCCATCC 436  
QY 509 GAACAGGGTGTCTGTGTAATAGACCAAGCTACTATCTACTCCGGAGAAACC 567  
Db 437 TGGACCTGCGCGGTGCTGGTGGTGGTGGAGCGGCTTACATGATTCAGCAGCGGCG 495

RESULT 13  
B0115402  
LOCUS B0115402 667 bp mRNA linear EST 22-JUL-2002  
DEFINITION EST600978 mixed potato tissues Solanum tuberosum cDNA clone STMDA85  
ACCESSION B0115402  
VERSION B0115402.2 GI:21916944  
KEYWORDS EST.  
SOURCE Solanum tuberosum  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE 1 (bases 1 to 667)  
AUTHORS Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C., Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and Karaycheva, S.A.  
TITLE Generation of a set of potato cDNA clones for microarray analyses  
JOURNAL Unpublished (2002)  
COMMENT On Apr 17, 2002 this sequence version replaced gi:20167364.  
Contact: Robin Buell  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Email: potatoc@tigr.org  
This clone is available through the Research Genetics, contact the Research Genetics for further information 1-800-711-6195 or cdna@resgen.com

Seq primer: T3.

FEATURES  
source Location/Qualifiers  
1..667  
/organism="Solanum tuberosum"  
/cultivar="Kennebec or Binjite"  
/db\_xref="taxon:4113"  
/clone\_lib="mixed potato tissues"  
/clone="STMDA85"  
/tissue\_type="mixed tissues"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; supplier: Combination of untreated and Phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes, tubers, or roots."  
178 a 133 c 160 g 196 t

BASE COUNT 178 a 133 c 160 g 196 t  
ORIGIN  
Query Match 4.4%; Score 47.2; DB 14; Length 667;  
Best Local Similarity 48.5%; Pred. No. 0.045;  
Matches 130; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 149 ACAATATCCAGACCCGAGCAAGAGATTAAGCGGTTCTTCGGGATTTTTCGGCG 208  
Db 399 ATATTTATCCCTGATCCTGAAAGCCGCTGCTTCGTCGAGCTCTTCTGCTGAAGATTCTGCC 458  
QY 209 TTAAGGAGAAATTTAGTTCTCGGTAACGGTTTCGGAGAACTCATATACCTCTCAA 268  
Db 459 TTGAATCTGAGTATATCTTCGAGGGTGGGTGAGATGAATGATTAATGA 518  
QY 269 TAGCTATAGTGAACTTTACATACCCGTTTACATACCTGTTCCACCTTTCCCATGTAGC 328  
Db 519 GATCGGTATGATCCTGCTGACAGATGTTGACTGCCACCCACCTTTCACATGATG 578  
QY 329 AGATAAGTGGCAAACTTCTCGGAAGACCCCTCGTAAGGTTCAACTGGAGAACTTTG 388  
Db 579 RAATTTGATGACGCTGTTAATGGAGCAGGTGTCATCAAGGTGCTAGGAACCCAGACTTTA 638  
QY 389 ATATAGACTTTAGAAAGATTTGAAATT 416  
Db 639 GCCTGGATGTAACAGGATTGCCGAAGT 666

RESULT 14  
AY105176  
LOCUS AY105176 1757 bp mRNA linear HTC 25-MAY-2002  
DEFINITION Zea mays PC0111982 mRNA sequence.  
ACCESSION AY105176  
VERSION AY105176.1 GI:21208254  
KEYWORDS HTC.  
SOURCE Zea mays.  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 1757)  
AUTHORS Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.  
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes  
JOURNAL Unpublished (2002)  
REFERENCE 2 (bases 1 to 1757)  
AUTHORS Coe, E.C.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA  
LOCATION/Qualifiers  
1..1757  
/organism="Zea mays"  
/db\_xref="MaizeDB:636533"  
/db\_xref="taxon:4577"  
/clone="PC0111982"  
/clone\_lib="Maize Mapping Project/DuPont Consensus



